



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 172388

TO: Tekchand Saidha
Location: rem/2A65/2C70
Art Unit: 1652
Monday, December 12, 2005

Case Serial Number: 10/621826

From: Edward Hart
Location: Biotech-Chem Library
REM-1A55
Phone: 571-272-2512

edward.hart@uspto.gov

Search Notes

Examiner Saidha,

Here are the results of the search you requested.

Please feel free to contact me if you have any questions.

Edward Hart

THIS PAGE BLANK (USPTO)

From: Saidha, Tekchand
Sent: Wednesday, November 23, 2005 10:11 AM
To: STIC-Biotech/ChemLib
Subject: 10621826 - sequence search request

RECEIVED
NOV 23 2005
STIC/BIOTECH/CHM LIB
(STIC)

10621826

Please search all commercial and US patent data bases (including interference files) for:

1. SEQ ID NO: 2 and
2. SEQ ID NO: 3

Thank you !

Tekchand Saidha
Primary Examiner
Art Unit 1652, Remsen Bldg. E02A65 (Office),
Remsen Bldg. 02C70 (Mail Box)
(571) 272 0940

November 23, 2006

Searcher: _____
Searcher Phone: _____
Date Searcher Picked up: 12/1/05
Date completed: 12/12/05
Searcher Prep Time: _____
Online Time: _____

Type of Search
NA# 1 AA# 1
S/L: _____ Oligomer: _____
Encode/Transl: _____
Structure #: _____ Text: _____
Inventor: _____ Litigation: _____

Vendors and cost where applicable
STN: _____
DIALOG: _____
QUESTEL/ORBIT: _____
LEXIS/NEXIS: DBH
SEQUENCE SYSTEM: DBH
WWW/Internet: _____
Other (Specify): DBH

THIS PAGE BLANK (USPTO)



STIC SEARCH RESULTS FEEDBACK FORM

Biotech-Chem Library

Questions about the scope or the results of the search? Contact *the searcher* or *contact:*

Mary Hale, Information Branch Supervisor
Remsen Bldg. 01 D86
571-272-2507

Voluntary Results Feedback Form

➤ I am an examiner in Workgroup: Example: 1610

➤ Relevant prior art **found**, search results used as follows:

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature
(journal articles, conference proceedings, new product announcements etc.)

➤ Relevant prior art **not found**:

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention.

Comments:

Drop off or send completed forms to STIC-Biotech-Chem Library Remsen Bldg.



THIS PAGE BLANK (USPTO)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 10, 2005, 11:46:00 ; Search time 9332 Seconds
(without alignments)
11268.794 Million cell updates/sec

Title: US-10-621-826-2

Perfect score: 1850

Sequence: 1 ccgagggcgccatccacgag.....tcagctcgatccgacgagg 1850

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_in.*

3: gb_env.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pr.*

9: gb_ro.*

10: gb_sts.*

11: gb_sv.*

12: gb_un.*

13: gb_vi.*

14: gb_htg.*

15: gb_pl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	256.4	13.9	22511	1	AF254925	AF254925 Streptomyces
C 2	242.4	13.1	85163	1	AY048670	AY048670 Streptomyces
C 3	242.4	13.1	292100	1	SC0939121	AL939121 Streptomyces
C 4	241	13.0	110000	1	AP006840_33	Continuation (34 o
C 5	241	13.0	110000	1	AP006840_34	Continuation (35 o
C 6	236.4	12.8	126933	1	DQ118863	DQ118863 Streptomyces
C 7	232.8	12.6	110000	1	BA000030_41	Continuation (42 o
C 8	231.6	12.5	110000	1	CP000010_06	Continuation (7 of
C 9	230	12.4	110000	1	BX571965_28	Continuation (29 o
C 10	227.2	12.3	110000	1	CP000031_23	Continuation (24 o
C 11	226.2	12.2	110000	1	AP006618_13	Continuation (14 o
C 12	223.8	12.1	2946	1	STMHUTH	M77841 Streptomyces
C 13	217	11.7	110000	1	BA000040_68	Continuation (69 o
C 14	216.8	11.7	49736	1	AF319998	AF319998 Stigmatel
C 15	215.4	11.6	1527	6	AR626602	AR626602 Sequence
C 16	215.4	11.6	6439	6	AR619418	AR619418 Sequence
C 17	214.6	11.6	110000	1	CR555306_34	Continuation (35 o
C 18	206.6	11.2	110000	1	BA000045_32	Continuation (33 o

C 19	198.6	10.7	203050	1	AL646071	AL646071 Ralstonia
C 20	191.8	10.4	301617	1	AE016911	AE016911 Chromobac
C 21	189	10.2	203050	1	AL646078	AL646078 Ralstonia
C 22	177.2	9.6	213732	1	AE001862	AE001862 Deinococc
C 23	176.8	9.6	10225	1	AE005048	AE005048 Halobacte
C 24	175	9.5	110000	1	CP000090_29	Continuation (30 o
C 25	173.4	9.4	10431	1	AE005773	AE005773 Caulobact
C 26	173.4	9.4	11099	1	AE004922	AE004922 Pseudomon
C 27	172.8	9.3	110000	1	AE014295_18	Continuation (19 o
C 28	172.8	9.3	349980	6	AX492785	AX492785 Sequence
C 29	172.8	9.3	349980	6	AX492786	AX492786 Sequence
C 30	172.8	9.3	349980	6	AX553952	AX553952 Sequence
C 31	172.8	9.3	349980	6	AX553953	AX553953 Sequence
C 32	171.4	9.3	11038	1	AE011796	AE011796 Xanthomon
C 33	171	9.2	110000	1	AE014292_09	Continuation (10 o
C 34	171	9.2	110000	1	AE017224_02	Continuation (3 of
C 35	170.8	9.2	10029	1	AE009324	AE009324 Agrobacte
C 36	170.8	9.2	13627	1	AE008290	AE008290 Agrobacte
C 37	170.8	9.2	110000	1	BA000012_68	Continuation (69 o
C 38	169.2	9.1	110000	1	CR555306_33	Continuation (34 o
C 39	164.2	8.9	110000	1	CP000009_12	Continuation (13 o
C 40	163.6	8.8	30390	6	CQ363732	CQ363732 Sequence
C 41	163.6	8.8	110000	1	AE017283_23	Continuation (24 o
C 42	163.2	8.8	1527	6	AR389675	AR389675 Sequence
C 43	161.8	8.7	1542	6	AX350741	AX350741 Sequence
C 44	161.6	8.7	9926	1	AE012259	AE012259 Xanthomon
C 45	161.6	8.7	110000	1	CP000050_31	Continuation (32 o

ALIGNMENTS

RESULT 1
AF254925/c
LOCUS AF254925 22511 bp DNA linear BCT 01-JUL-2002
DEFINITION Streptomyces maritimus enterocin biosynthetic gene cluster,
complete sequence.
ACCESSION AF254925
VERSION AF254925.1 GI:8926181
KEYWORDS Streptomyces maritimus
SOURCE Streptomyces maritimus
ORGANISM Streptomyces maritimus
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomycetes.
REFERENCE Piel, J., Hoang, K. and Moore, B.S.
AUTHORS Piel, J., Hoang, K. and Moore, B.S.
TITLE Natural Metabolic Diversity Encoded by the Enterocin Biosynthesis
Gene Cluster
J. Am. Chem. Soc. 122, 5415-5416 (2000)
REFERENCE 2 (bases 1 to 22511)
AUTHORS Piel, J., Hertweck, C., Shipley, P.R., Hunt, D.M., Newman, M.S. and
Moore, B.S.
TITLE Cloning, sequencing and analysis of the enterocin biosynthesis gene
cluster from the marine isolate 'Streptomyces maritimus': evidence
for the derailment of an aromatic polyketide synthase
Chem. Biol. 7 (12), 943-955 (2000)
REFERENCE 3 (bases 1 to 22511)
AUTHORS Hertweck, C., Jarvis, A.P., Xiang, L., Moore, B.S. and Oldham, N.J.
TITLE A mechanism of benzoic acid biosynthesis in plants and bacteria
that mirrors fatty acid beta-oxidation
Chembiochem 2 (10), 784-786 (2001)
REFERENCE 4 (bases 7073 to 7798; 12734 to 14128)
AUTHORS Xiang, L., Kalaizis, J.A., Nilsen, G., Chen, L. and Moore, B.S.
TITLE Mutational analysis of the enterocin favorskii biosynthetic
rearrangement
Org. Lett. 4 (6), 957-960 (2002)
REFERENCE 5 (bases 1 to 22511)
AUTHORS Moore, B.S., Piel, J. and Shock, D.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) Department of Pharmacology and Toxicology,

FEATURES
source
College of Pharmacy, University of Arizona, Box 210207, Tucson, AZ
85721-0207, USA
Location/Qualifiers
1. 22511
/organism="Streptomyces maritimus"
/mol_type="genomic DNA"
/db_xref="taxon:115828"
513. .1004
/gene="encE"
513. .1004
/gene="encE"
/codon_start=1
/transl_table=11
/product="putative regulatory protein EncE"
/protein_id="AAF81720.1"
/db_xref="GI:8926182"
/translation="MAQRMMQIGEVAERTGLSRTIRHYEEVGLIVPSARSKGGFPL
YTQADVSLMIRMKPLDLSLEMRDLLEITDRLGQDDPPAPVERERLRERDAYR
KVADARCBETLRARLAMEDFAATWRDGMNADARSASERATSGGLRLRLWLTSDAPG
LAG"
1749. .2600
/gene="encF"
1749. .2600
/gene="encF"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator EncF"
/protein_id="AAF81721.1"
/db_xref="GI:8926183"
/translation="MDVEVLGHVTLKVGDERRTSPAKARQVLALLLVNANRMVSTDV
LIEELWAPRPSALSALTQVYFOLLRIKRPDECPKOLLVTPNGYFRVPPENV
DLQYERHLAGAQQSGENNAAVEQFRAALHIFAAPPLANVRGGMHLEAEIYRLNE
SRIGQVETRIAELELGRHVEVLSGLASQCQHEGFRHQLMALRYCGRRRQALK
VFNRQLSVTELGLPEPSRELQMLQALLSADPRDLTAERTFFLRTTADPAPLQVWP
AASPWKV"
complement (2672. .3163)
/gene="encG"
complement (2672. .3163)
/gene="encG"
/codon_start=1
/transl_table=11
/product="EncG"
/protein_id="AAF81722.1"
/db_xref="GI:8926184"
/translation="MTTAGADPTQAESALTAQPKPEETAALFAEIQQFYGRHWRAM
DEGRVEDWTGPFMDAPFATNARPEPQGRAEIARNAEAAARQLOEKGILRRHCVTTL
ELQQAGLTLAKTYALITKTVVGSRSELEFVFCEDVLVRNEGRWFIHRQVPRDDL
PKR"
3330. .4937
/gene="encH"
3330. 4937
/gene="encH"
/codon_start=1
/transl_table=11
/product="putative acyl-CoA ligase EncH"
/protein_id="AAF81723.1"
/db_xref="GI:8926185"
/translation="MEPAQAPTPQVDARPEAGPRWETIGELIKDAALLHGKKEFL
RCGDKTUSFSTDTTRDLAQALIAQGVKGDVAVVMDNDVPLSWFAAIAKAGAIT
VFNVTRFGATLAHLKDSQAVRVLASPGCVPLARDVSGSIGCHPCAVRTLRELEADF
DRPVDGCLSAHADDVNFQYTSGTTPGPKACMLSHDYWLRTANWIAVHSLRPDDVV
LTAQAFSYMDQKAVMCLMGVPLVLPRESASGFWHSVQHRATLTYYLVSWPMLL
YKQPHSGDRDHMRVLVCSGIPRDLHAFEDRNGAARVEYVSGSTGLDLIMPGE
ATVGSAGYPPSGREVIVADERQRPVQPGIGELVGRPPMMKYNWNNPSTERA
GGWYHTDGLBRAAGSVYHAGRLKDMIRGGENIAAAEVESVLEAHPAVLAAALAGI
PDELFLGPKAFQLRPGYRPTTARSVLATRRHLAKFKVPAYVBFVDSFMTPSA
RIQKQLLRPGDDQRTGTGAFAADAWA"
4934. 5710
/gene="encI"
4934. .5710
/gene="encI"
/codon_start=1
/transl_table=11

/product="putative enoyl-CoA hydratase EncI"
/protein_id="AAF81724.1"
/db_xref="GI:8926186"
/translation="MSDVQQLSSADLTVTVDSVAVIAMSRAKLNALTRGMRDLA
ELLRHYGDGSGARGVITGQAKAFSAGMDREVAGPEVDLLAEMLFNDITRAALITE
VPVVAALGIAGVGAGEMALSFDARIATAQAAISWPEVTGLTVSNAAISLPLRLAGA
PRALHLTMNSARLSAQALGLDDIVEPDDLIIPAAIRLIHQWTPPSSSTAHLRLM
RPSLRVAAAMAREGAARQVQESGIAQAQIATVLAGHEK"
complement (5856. .7058)
/gene="encJ"
complement (5856. .7058)
/gene="encJ"
/codon_start=1
/transl_table=11
/product="putative beta-oxoacyl-CoA thiolase EncJ"
/protein_id="AAF81725.1"
/db_xref="GI:8926187"
/translation="MTSGATYRDAAILPLTALWTSPPFARMGGALSHLSLDLAAVATGR
ALSERGIDRTIEGMVLGWTVPQPAIFYGATITLAARLGMFQVSGPMLNQACATVAAL
HQAAAVGTGVPOLI VTTDRTNSGPDLSWPSAGSEQPVRENWVRDSSFQDDPVGTQ
SMLTAEVAVAECGFTRDDLDALAVLSEQYTAALSDRTPORDPFWGIEIEQDGLGT
VRLLSDEGVPTVADSAGQLPTVRPGSVHTAATQTHEPADGTGALVITTTFCARFPADR
GPVRLVASGFARVEPARMPRALVPAAQAALHAAGLKIDTMDALSTNHPFAVNDLIYA
RETGVTLAEMNASGCSLLFGHPQPGTGLRTVTELAHLVLRKAGGGLGLFTGCAAGDTAA
ALIIIEVTD"
complement (7073. .7798)
/gene="encK"
complement (7073. .7798)
/gene="encK"
/codon_start=1
/transl_table=11
/product="putative O-methyl transferase EncK"
/protein_id="AAF81726.1"
/db_xref="GI:8926188"
/translation="MTIDASPESAFEGEQSYEHRRLRYDAVVVHASNRLWRCR
KSRLLHEYNDWITSRHLEVGSGYLDHCRFPAGTPELTLLDLPFLKFAAHLRR
YTPRCVQADLLEPLDPFGSGFTSIAYNYVMHGLPEPPGGKQIVFKHLRALADPGLV
FGSTVLSQGVKHTALSRRNILYQRQGSFHNQDHSVHGLHQALEEHFASHWIEVRGSV
ALFAARASQTTAPRHPCASLDQR"
complement (7826. .8626)
/gene="encD"
complement (7826. .8626)
/gene="encD"
/codon_start=1
/transl_table=11
/product="putative ketoreductase EncD"
/protein_id="AAF81727.1"
/db_xref="GI:8926189"
/translation="MPAHTSPSVALVTGTSIGLISIAQALLQDGLKVIITGRDPGK
LETVRSKLWEDDQFDVAGYTCVDRDRDAVHDMVERCAKEHGPVKLVNNGRSGGV
TAAIPDELWLDVINTNLNSVFWTVREVIRSSDAIKOPAGRIINIASYKGGQGVPGAP
YSAKAGVIGFTKALAKEFAAHGGPTVNAVCPGYVETPMAQDIRASAKLNNVTEQV
LADFEAKIPUGRYCTPEEVAGWVRVLYSPAAASVTAQALNVCGGLGSI"
8815. .10086
/gene="encA"
8815. .10086
/gene="encA"
/codon_start=1
/transl_table=11
/product="putative keto synthase alpha EncA"
/protein_id="AAF81728.1"
/db_xref="GI:8926190"
/translation="MSRHVITGLSVLAPGSTDTGEGFKMITAGRSAIRRTAFDPTP
FRSQAGEVDLDPLACGFSRREVRRLDRASILLAVACARRAAVEAGITSGSISDPRIG
VSVGNVGSATSIENEYVLSDEGRQWLVDQKQSPHLFDYFVPGSARETAAWAGAE
GPVSVISAGPTSGTIGDALGHCLOIQESADVMVAGATDAPVTPIAVACFAIKATRR
NDDEPRFPFRDTRDGFALAEGAAMLVLEDSHARRRGADYGTITGYAAHSNAYHM
TGLSRDGNEMARAITTALDEARPTAVGVINAHGSGTKQNDLHETAEFAKALGEHAY
NVPVSSIKSMIGHSLGAIGAEVATCALAENGWVPTANLHEDPDCCDLDDYVPLEAR
DVLDDVVLTVGSGFGGFSQAWIVEKWRGGHV"
10083. .11306
/gene="encB"
10083. .11306

```
/gene="encB"
/codon_start=1
/transl_table=11
/product="putative keto synthase beta EncB"
/protein_id="AAF81729.1"
/db_xref="GI:8926191"
/tranlation="MSDIDIVTGMVTAPTGLSVKEYNATLACTQGTAPLOEFDPTP
YTRLAGQITGFEAKRHLPRLLPQTRVTRALVAADRALEDQNDVDELGTVECGV
VTSNATGGFESHREMRQWTOGPRVSVYQCFWYAVNTGQISIRHKMRPGVGVV
AEQAGLDIAIGHARTTIRGRSLMVTGVSFDPMGVSHLASGTVSTATDPATAY
LPFQACQVYVGGGALLVMEDAQSARGRTDRPYKITYAAATFDPAPCAQPTMGL
RRAVDLADLAGLKPDAIDLVAADAGVADLDAOEADALRAVFPYGVPCAPKTMVG
RLMAGGSPDVVCCALLALRDNVIPAVNVTVDDYQIDLDVDEPRTCLPLRALVLR
GAGFNSAVIVEHTP"
11354..11602

Query Match      13.9%; Score 256.4; DB 1; Length 22511;
Best Local Similarity 51.4%; Pred. No. 4,1e-24;
Matches 726; Conservative 0; Mismatches 666; Indels 20; Gaps 5;

Qy 183 GCCGTGAGCTGGATCGCAATCATGATCTGGACACAGGCCCATGCGGTGGAGCGCGGC 242
Db 17831 GTCATAGAGCTCGACATGAACGTCAAGCTCGACCAACTTGAGGACGCGCGCGACAGCGC 17772

Qy 243 GCGCGATTGTCTTGCCTCCCGCGCGGACCGGTGCGGTGCGTCCGAGCGCGGCTC 302
Db 17771 ACGCCCGTGGAGCTGTCCGACCCCGTCCCGCTCCCGCGCTCCGCGGACGTGTG 17712

Qy 303 GCGCGTGTCACTCCGCGAGCGCGCATGTCTACGGACTGACAAACCGGCTTCGGTCCCTT 362
Db 17711 GTGAAGTTCGTGACGACGACGTGTATCTACGGGTCAACACAGCATGGGGGCTTC 17652

Qy 363 GCGAACCCGCTGATCTCAGGTGAGAAATGTCCGAACTGCGAGCGCAATCTTGTTCATCAT 422
Db 17651 GTCGACCACTCGTCCCGGTGTCCAGCGCGCGAGCTCCAGGAGAACTGATCAACGCG 17592

Qy 423 CTGGCAGCGCGGTGGACCGGTGCTTACTTGGACGCGCGCGCGCATGGTGTCTGGCG 482
Db 17591 GTCGCCCAACAGTGGGGCGGTATCTGGACGACACGACCGCCCGGACCATCATGTGTGCC 17532

Qy 483 CGTCTGGTGTGATCGCTCAGGAGCCTCCGCTGCCAGGAGGACCATCGTCCGCTG 542
Db 17531 CGCATGHTGTGCTGGCGCGGGGACCTCCGGATCACCCCGCGGATCTGGACAGCTG 17472

Qy 543 ATCGACCTGTCAATTCGAGCTCGCTCCCGCGGTTCCTCCAGCGCGGACCGTGGCGCG 602
Db 17471 GTGGCGGTACTCAACCGCGGATCGTGCCTGATCCCGGAGAGGGCTCTTTGGGCACC 17412

Qy 603 TCGGTGACCTGACACCGCTTGGCATATGTGCTCTGCTCCAGGCGCGGAGACTTC 662
Db 17411 AGCGGTGACCTCGGCGCGCTGGCGCGGATCGCCCTTGGTGTGGCGGGGACGTGGAGGCC 17352

Qy 663 CTGGACCGGACGAGACCGGCTTGACGCGCGAGAAAGGCTCCGCGCGCGGACGCGG -TGCA 721
Db 17351 CGCTAC---AACGTCAGATCATGCCCGGGCGGAGAGCCCTGTCCGAGCGCGGCTGCA 17296

Qy 722 ACCGCTCGATCTCTCCATTCGATGCACTGCGGTGTGTCGTAACGAGGACCTCGCCATGAC 781
Db 17295 GCCGATGAGCTGAGCTACAGGATGGCTGGCCCTGTATCAACGCGACGTCAGGCATGGT 17236

Qy 782 CGGATCGCGTGTGATGTCACGCTGCGGCATCTCGGCACTGGCGGTGGCGCT 841
Db 17235 CGGCTGGGCACCATGTTCTCCAGCGCGCGGCTCGTGGACCGCTACCTACGAGT 17176

Qy 842 GACGCGCTGTGCGGAATGTCTGAGAGCGCGGACCGAGGATCGGCGCGGCGGCACTGTC 901
Db 17175 GTCGCGTGTGCTGAGGCGCTGGCAGGATGACGAAACGTTTCGACCTTCGCGTGA 17116

Qy 902 CGACTGCGCGCGCATCCCGGACAGAGGACCGCGAGCGAGGTGCGCGCGCGCGTGA 961
Db 17115 CGCGGTCAAGCGGACCGCGGACGCGTCAAGTGGCTCGCGGTGTGGAGGGGCTTGC 17056

Qy 962 CGGACGCGCGGCGGTGTGTCGCGACGTCATTTGCCAGCGGAGGCTCGAC-----GC 1012
```

```
Db 17055 CGATCGCACTGGCGGTCAACGAACTGGAACCCAGGACAGACCTTGGCGGAGAGATGGG 16996
Qy 1013 CGGCGATATCGGAGCGGAGCGGAGGAGGATCCCTACACCTTGCCTGCGCTCC 1072
Db 16995 CACGTCGCCAAGCGCGGTTCGTCGCGATCGAGGACGCTTACTCATCCGTCGACGCC 16936
Qy 1073 CGAGTTCTCGGGGGGGCTTCGACACGCTCGCATGGCATGACCGGTGCTGACAGTGA 1132
Db 16935 GCAGATCTCGGTCCGCTCGGTTCGATGTGTGACCGGATCGGGGCGACCTCGACGAGCA 16876
Qy 1133 GCTGAACGCGGTGACCGCAATCCGCTGTTTCGCGCGCATGGACGCGTCCCGCCCTGCA 1192
Db 16875 GCTGAATCTCTCAACGACCAACCGATCGTCTGCGGAGG---AGGCGGAGGTGTTCA 16819
Qy 1193 CGGGGGCAATTTTCATGGGCCAGCATGTGGCGCTGACGTCCGATGCGTCCGACGCGCT 1252
Db 16818 CAACGGGCACTTCACGCGCAGTAGTGGCCATGGCCATGGACACCTTGAACATGGCCCT 16759
Qy 1253 CACGTTCTCGGGGGCTTTCGAGGCGCCAGATTCACTGTCAGACAGTGAAGGCTGAA 1312
Db 16758 GGCACCGGTGACCAATCTCGCCAAACCGCGCGTGGACCGCTTCCTGGACAAAGCAACAG 16699
Qy 1313 CGTGGGCTGCGCCCTTCCTCCACCGGGGCGCGCGGTTGAATTCGGGCTTCATGGG 1372
Db 16698 CAACGGGCTGCGCGCTTCCTGTGCGGGAAGATCGGGAATGCGCTGGGCTGATGGG 16639
Qy 1373 CGCACAGTGAAGCGGACCGCGCTCTTGGCCGAGATGCGAGCC---ACGGGACCTGCTTC 1429
Db 16638 CGGCAGTTGATGATCGCGCTCGATCACCGCGGAGACCGGACCTTCCGATTCGATGTC 16579
Qy 1430 GATCATTCGATCTCCAGAACCGCCCAATCAGATGTGCTGCTGGGACCATCGC 1489
Db 16578 GGTGAGTCCCTCAAGTAGTACGCGGAGCTTCAGGACATCGTGTCTTCGGAATCGTCCG 16519
Qy 1490 CGCGCGCTCTGCGCGGAGAGATCGACCTTGGCGGAGATCTTCGGATCTCGCTCT 1549
Db 16518 CGCGCGCGCGCGCGGAGGACTTACCACACGCTGCTACGTTGGTGGCTTCGAGTGT 16459
Qy 1550 CTGTTTTCACAAGTCCGAGCTGCGTGGCG 1581
Db 16458 GTGCGCTGCCAGGCGCGTCCGACATCCGCGGCG 16427

RESULT 2
AY048670/c
LOCUS AY048670 85163 bp DNA linear BCT 05-NOV-2002
DEFINITION Streptomyces globisporus enediyne antitumor antibiotic C-1027
biosynthetic gene cluster, complete sequence.
ACCESSION AY048670
VERSION AY048670.1 GI:24575076
KEYWORDS
SOURCE
ORGANISM
Streptomyces globisporus
Streptomyces globisporus
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE
1. (bases 1 to 85163)
AUTHORS Liu, W., Christenson, S.D., Standage, S. and Shen, B.
TITLE Biosynthesis of the enediyne antitumor antibiotic C-1027
JOURNAL Science 297 (5584), 1170-1173 (2002)
PUBMED 12183628
REFERENCE
2. (bases 1 to 85163)
AUTHORS Shen, B., Christenson, S.D., Liu, W. and Standage, S.
TITLE Direct Submision
JOURNAL Submitted (27-JUL-2001) Chemistry, University of California, Davis,
One Shields Ave, Davis, CA 95616, USA
FEATURES
Location/Qualifiers
1..85163
/organism="Streptomyces globisporus"
/mol_type="genomic DNA"
/strains="C-1027"
/db_xref="taxon:1908"
complement(8..658)
source
CDS
```

/note="ORF (-7)"
 /codon_start=1
 /transl_table=11
 /product="hydroxylase-like protein"
 /protein_id="AAL06648.1"
 /db_xref="GI:24575077"
 /translation="MGMTGARPRVLVVGAGLAGTATAIRLLHARRRPLEVVLERRAA
 YRAGAVYHRDGNFVNIQAGMSVFREDVLDFINWANOEARDRDWRPRWASWKF
 TEOGPAPRRIQDYLDARLVEAARESCPVVLVEADGEALDASHDPCFSTVVRGLTP
 YLEGLRPGPLDADHVLATGLELKEPPFAAGVAGHSFVRNPISAPGIL"
 complement(930..1478)
 /note="ORF (-6)"
 /codon_start=1
 /transl_table=11
 /product="viral infectivity potentiator-like protein"
 /protein_id="AAL06649.1"
 /db_xref="GI:24575078"
 /translation="MPRIPTHITDPAAGDILRLLEKFRGRLVNIHGMASPVVL
 ETYAATGVAEHTGFQARTREAIALAVGADVACAYCOAAHTVSAKVAGFTLEETVAI
 RRGTPGDVLEALVQVAREIAGEVGEASDASWNAVAQGWTDTELAEVFVHAVNLY
 TNYFNHARTEIDVPGVPDIDS"
 complement(1649..2713)
 /note="Orf (-5)"
 /codon_start=1
 /transl_table=11
 /product="truncated methionine synthase-like protein"
 /protein_id="AAL06650.1"
 /db_xref="GI:24575079"
 /translation="MTIATEPIGSLPRSATLLHALAAHQAQGTLDATDLAQOQAVAD
 TLRLERVSFVLVDGSKSPFATYVGLDLSPDGAVIPFADGHTROLPSITQGP
 FRYQRAETYLRAARQITDRFLQAVIAPALSALSIYPATPIEGYPRQFLRLADAE
 ADRLCGDAGAVVQIDTFRGLSKLDPGGLDDDFIALNNEVLGRFSAETARIGV
 HPCPGQDQSTHSDIDYAEALLPKLFOLKAGNFYLEAGADPERVLSIVRDHLPPAA
 RVGLVDTPIDVETPEVRDRILLAAKYPVEQLGTCDCCGFSFPADTSTTRDLA
 FAKIEARVGTALAEVILG"
 complement(2851..3237)
 /note="ORF (-4); similar to SCP1"
 /codon_start=1
 /transl_table=11
 /product="transcription factor"
 /protein_id="AAL06651.1"
 /db_xref="GI:24575080"
 /translation="MSSLATDSGLVHEAKVWVLEDPGLDYVQALDKTPRRKNKRY
 ARDGRMIGYTELGADABDPSGLYRRVFLLPHDRDSDEPGYVQAGPGEAVDPRT
 IEPNVGEKTPRQLGTSSTVAATGS"
 complement(3442..4971)
 /note="ORF (-3); similar to SCP1"
 /codon_start=1
 /transl_table=11
 /product="putative primase"
 /protein_id="AAL06652.1"
 /db_xref="GI:24575081"
 /translation="MSSASGPRPDATAAQOMALETTTAPAPLIPVQAAPAAVAA
 TAWAGEIPASLDTRGNAGLFLVRLYRDQFRHVEGLGWSYDGYRWKRAGGEKAAALWAA
 GEMAEENPGSDPRLGFLDRELLHHKRTLTSTGMKALLTQAKASPOLSLDPTLDGDP
 YALCTPGVDLGRNMRKDPDTRDHSRATSASPODIPTPRWRHFLIEDTFGSDAER
 EMIDFLLHGGYSTGNGAQLPFLHGCKNGKSVLLDVMYIOLGDYADAAPPGTLM
 DRAGYSHSTELTGHRRLLVCSLKPNDRPFDEARVRLITGGDKIKARMRDYRSF
 TPYTHKLLGNHPEVSTGFGFWRRLIRLPPFRIVDERKIDNLNAVELVQDSGPGIL
 HWLTGARYLATRDTLAGDVRRIATSAVANTEDHIGRFLEACCLHDPENSLRVEQ
 GLLYTSYSTWCAHSEGIPTGATAFATRVQEVGLASPADMKSNGRKFYPNLALAA
 E"
 5982..7478
 /note="ORF (-2)"
 /codon_start=1
 /transl_table=11
 /product="glycerol phosphate ABC transporter"
 /protein_id="AAL06653.1"
 /db_xref="GI:24575082"
 /translation="MRALPSGGLRMAGDAPTAAGNRCPPTPIQRPLILRRQTLQAP
 EAKNGSPQQRERIVLDVWLANYPPTDYGRDFFLAPLRERAAEFERAPRYRVDIN
 GHDFWTIPEKVATAREGPHPIAGYATDSQLARDARPDGKPVFTSVBAALGRTE
 ILGHPVVVEDLDPVVRDSYFGGELVSLPLTVTMTLCYANSSLLARAGVPELPRTWDE

VEAAQAVASVDGPGHGHTIANDGWVFOQAVALONGVLTQDNNGRSGSATTVDTVTS
 EMLDVRWTHLHERGHLYTGQPSDWGFAFEAVQOKVAFPTDSSKAAARELIIQAGAO
 AGEFVAVFPLPRNAKAPVQPSGSLWLAAGLDGTTDGLLALTYLISPANAADW
 HRTNGFVPTGAAGELLEATGFWDRRPOQRVAGEQLKASDRSPAALGALLGDFPAVNE
 VITTAAMDVLRSAGADPAFAEAGVAAQQLLDAYNARNRSGSGTSPSAV"
 complement(7573..9927)
 /note="ORF (-1)"
 /codon_start=1
 /transl_table=11
 /product="uvrA-like drug resistance pump"
 /protein_id="AAL06654.1"
 /db_xref="GI:24575083"
 /translation="MYGSOEENSVSVDTPTAGAKTSGPGDGIRIAGARIHNLKDVSL
 TIIRNOITVFTGVGSGKSIIVFDIVAVEAQROLNSTFSWIRNQLPKYERPOAEATE
 NLPRTVIVQKPVGNARSTVGTMTDVPQIMRALFARFEGTTKDDGSSLGVSASFND
 PQMGFCDCGLGSIADLDKMLDRSKSLDDGAVLPEYKVGSPDQWQWAKSGRLDPA
 KPIAYSAEELDTLLRGTKGKTLTKTSFTQNYEGIADRFEKLNKRLDLSALSDRK
 RIERVFVTDGVCGRGARNAAALDVRIDGKNIADYSMEVRDLTEVLAVTEPAA
 APIAKAARTALERIVSIGLYLTDRLPTADLSGEGQLKWRHLGSLAGLTYIFDE
 PSILGHPDVGRILNLRALDKGNTVLVVEHDPDVIETADHIVDVGPGAGVHGEIV
 FEGSFALKRKAKTRTGEARRGARVKEEVRPPTGELTVENADLHNLKQVSVAVPTGVL
 TAYTVAGSGKSLISGAFMEAHDPDAVVDQSAALASSRSTPVSYLGLMPLRLKPAK
 ETGANASLFSNKGSCBCECCQGRGVIIITELAFADPDVTHCGVCEGRAPKEVLEHQLR
 GSIADVLELPDAVEAFFTEKALLPKLRALVDGLYLSGLQPSLSGSGELQRIKL
 ADQLHRTGVYVLDDEPTTGLHMDVDVTLKLDLGLVEAGNTVVVIEHNLVDVQADWI
 IDLGPDGREGGEIVFTGTPKDLLAADSSLTGTYLRRHLKAQAG"
 complement(9982..11349)
 /note="ORF0"
 /codon_start=1
 /transl_table=11
 /product="Na/H efflux pump"
 /protein_id="AAL06655.1"
 /db_xref="GI:24575084"
 /translation="MAVLATSPIGCEPLTVFLLOVCLLILACAYGLGRGLTRGLPLP
 VGLSTAGVLGPTLLQGISPLSGRFPEDISQAHLDDAFQCQFGLLLVAIAGAQFDP
 RILRKGLGAARVSLAGLITPLGLGIATGYLVASLILADSGERGVSFALFLGVAMCVTA
 LPVIAKTLADNLNTHRNQGLIIAAAVFDDAVGMLLIALVTALASGAAGGVPVLTWTA
 WTTVFVAAACAVGSPIGRRLSRGDSRPVSAVTGVVAVVLYGALTAAAGMEALFGA
 FVAGATLLRHIAVRLAPLRTLMAYAPAPVFLSGVGLMDLTALAEPSVLLTGLGVLL
 VATGFPAGYVAARSGMSRYEGLAGAGNRMGMEVIVIALVGLRIGVLDVTFTFI
 IIVLALITSVTAPILRWASSRIVLEEDETERGDRLAGMWTPEALSGGPAPKASREBK
 TPDTS"
 complement(11351..12835)
 /note="ORF11"
 /codon_start=1
 /transl_table=11
 /product="alkylhalidase"
 /protein_id="AAL06656.1"
 /db_xref="GI:24575085"
 /translation="MDVSAQYDVIYVGGPGAGSTVSTLVRKGRHVLQLEKETPPRYO
 IGESLAPSTVHGIAHLGLVSDCLKAAFTIKHGGTFKMGANPEPWTFTDFAVSRMPGA
 TGYAOVERMKFPQIILLNARRHGVRENSDVLVDLKAEDGRVGRVYRDSREHRE
 VGRFVVDASGNTGGYKKGAKREYSPPFRNLALFGYFNGKRLPKPKNSEIFCTIF
 EHGWFYIPLSPELTSVGVAVNRDSASLVQGDPEKAMESFIAACPLIAEKLSEARTV
 EGYGBELRYKDWYSNTKFWPAGMALVGDACFVDFVSSGVHLATYSGLLSARSLN
 SCLDGSVDTSAPEEFEARYRLGYFYEFLVGFDYDMHSEDSYFWOARKISNTADS
 LOSFVELIGVSDDEPADTARAKORFSEASDEMAGMTHTTAAGDADGPHGAVPGKVGR
 LMQSAALQARALLGDEAGNERPVROGGLAASDTGLGWVFGARS"
 13012..14079
 /note="ORF10"
 /codon_start=1
 /transl_table=11
 /product="dntp-glucose synthase"
 /protein_id="AAL06657.1"
 /db_xref="GI:24575086"
 /translation="MKALVLSGGSGTTLRPIYSAMPKQLVPFIAGKPVLEVLDNIRNL
 DIKEVAIVGDWQAQEIIEAMGDGSRFGLRTYIRQOPLGIAHCKVLARFLDDEDFV
 LYLGDIMLDLSAQAGHFLHTRPAARIIVRVQDPDRAFPGLDGEGRVLRVLEKPR
 EPDSILAAVGVYFETADVHEAVDAIISPRRGELEITDAIOMLLEOGLPVBAGRYTDVW
 KDTGRVEDVECNRMRLQALQVSGVEDPESELVGAVVVEGARVRSVRSVGPVAVIG
 AGTVSDSQIGPYASIGRRCITVRASRUSDIVLDDASILAVSGLHSLGRGARIAPG
 ARGEARHRLVVGDDHQIETAA"
 complement(14212..14643)

```
/note="Caga"
/codon_start=1
/transl_table=11
/product="C-1027 apoprotein"
/protein_id="AAL06658.1"
/db_xref="GI:24575087"
/translation="MSLRHMSRRSRFGVWVASIGLAAAQSVAFAPAFSPVSPASG
LSDGQSVSVSGVAGACTYTAOCAPVGGDCCNCPATATSTFTDASGAASFVVRK
SYTGSTPGTTPVGSVDCATACNLGAGNSGLDLGHVATFG"
CDS
complement(14690..15922)
/note="ORF8"
/codon_start=1
/transl_table=11
/product="amino transferase"
/protein_id="AAL06659.1"
/db_xref="GI:24575088"
/translation="MSGGPFRLDSLPTFPFSEPLHVRPNVRSRDLRMLERIDGALERL
WFTNDGPIVRFEPVAVELTOVRHCVAVSNAATGIQVAALKALGIPGDEIVPQFTWV
ATAHALDWGAPVPVCELDDEGTADVAVHVERLIGPRTRAILDVHVRGRPARIDELTK
LAAEGLHLFPDAHAFGCTYRSKPIGFGFTAEIIFSFOATKFNVSFEGLAVTDDDL
ADRLRMRHQGLNAAHEITGSGTVARMHEISAAAGLTSLEADHPTAINRRNRYIEQ
YLDGLPVRVQPDNELNSCOYVIEVDVAVRAGLRDELQAVLQRHNVLARVYFSPG
CHSCEPYRSLARLAPDLPKVEALTEVLSTPTAVGPBEVRGVCRILRAAVDGA
VPEIHSTEGDAGGRPAR"
complement(15919..16653)
/note="ORF7"
/codon_start=1

Query Match      13.1%; Score 242.4; DB 1; Length 85163;
Best Local Similarity 50.3%; Pred. No. 1.8e-22;
Matches 749; Conservative 0; Mismatches 706; Indels 33; Gaps 5;

QY 227 CGTGGCGAGCGCGCGCGGATGCTTGTCCCTCCGCGCGCGGACCGGTGCGTGC 286
DB 40901 CGTGGCGAGGAACCGCGCGGAGTGCAGGTACCGCGCGGATCCATCGCGAAGGCCAGAA 40842

QY 287 GTCGAAGCGCGCTCGCGCTGTCTATCGGAGGCGCGCATGCTACGGACTGACAA 346
DB 40841 GAGCGGAGATCTTCGAAGGATCGCGAAGAGATCCCATACCGGAGGAGGAGGAGG 40782

QY 347 CGGCTTCGGTCCCTTCGGAACCGCTGATCTCAGGTGAGAAATGTCGAACGCTGCAGGC 406
DB 40781 CGGTTACGCGAGATGATCTATCATGTCAGTGCAGAAAGTCGAAGTGCAGACTGCAGAC 40722

QY 407 CAATCTTGTCATCATCTGGCGAGCGCGGTGGAGACCGGTGCTTGACTGGAACGCGCG 466
DB 40721 CAATCTGCTGAGTACCAAGCGCGGAGTGGTCCGCTGTTCGCGGAGGAGCGAGCGCG 40662

QY 467 CGCATGTTCTGGCGCTGTGTGTCATCGCTCAGGAGGCTCCGTCGCGAGGAGG 526
DB 40661 GCGGATGTCGCGCGCGCTGGAACACCTTCGCAAGGCGCACTCCGCGGTGCGGCCCAT 40602

QY 527 GACCATCGCTCGCTGATCGACCTGCTCAATTCGAGCTCGCTCGGCGCTTCCAGCGC 586
DB 40601 CATCTCGAAGCCTCGCGAGTACCTGAAAGGAGGATGACCCGCGCATACCGGAGAT 40542

QY 587 CGGCACGGTGGCGCGTGGGTGACCTGACACCGCTTCGCGATATGFGTCTTCGCTCCA 646
DB 40541 CGGCTCACTCGGCGGAGCGGACCTGGCTCCCTCTCCACAGTTCGCGAGCACCTCAT 40482

QY 647 GGGCGGGGAGACTCTCGAGCGGAGCGGACGCGGCTTGACGGCGAGAGGCTCCG 706
DB 40481 CGGAGAGGGTACGTCTCGT---CGGAGCGAGCGCGGTGGAGACCGCCAGGTGCTGGC 40425

QY 707 GCGCGGAGCGGTGCAACCGCTCGATCTCTCCATCGGATGACCTGCGGCTGCTCAACGG 766
DB 40424 CGAGGGGGGATCGAGCGCTCGAACTCGGCTTCAAGAGGGGCTTCGCACTGATCAACGG 40365

QY 767 GACCTCGCCATGACCGGGATCGCGCTGGTGAATGCTCACGCCCTGCCGCCATCTCGGCAA 826
DB 40364 CACGTCCGGGATGACCGGTCTGGCTCTCCCTGCTGTCGAGCGGCGCTGGAGGAGGCCA 40305

QY 827 CTGGCGGTGGCGTTGACGGCCCTGCTTGGCGAATGCTGAGAGGCGGACCGGAGCATG 886
```

```
DB 40304 CGAGCCGAGATCGTGCAGGCTCTCTCTCATGAGGCGGTACCGGATGCACGAGCCCTT 40245
QY 887 GGCCCGCGCACTGTCTCGACCT---GCGGCCGCATCCCGGACAGAAAGAGCGCGCAGCGAG 943
DB 40244 CTTGCGGAGGCGGACGACATAGCCCGCGCACGAGGCGCAGATCGACACCGCGCCAA 40185
QY 944 CTTGCGGCGCGCGTGGAGCGGAGCGCGCGGTGTCTGGGACAGTCTCATTCGCGAGCGGAG 1003
DB 40184 CATGCGGCGCTGTATGCGGCGGACGCGGACTGACGCTCGAGCAGCGCGACTCTGCGCGGAGA 40125
QY 1004 GCTC-----GAGCGCGGATATCGGACGAGCGCGGAGCGCGGCGGCGAGCA 1048
DB 40124 ACTCAGAGGACAAAGGAGGCGCGGACGAGAGCTCAGCGCTCGGAGATCTACCTCGAGNA 40065
QY 1049 TGCCTACAGCTGCGCTGCGCTCGCAGGTTCTCGGCGCGGCTTCGACACGCTCGCATG 1108
DB 40064 GGCCTACTCTGCGGCGCATCCCGCAGGTCTGCGGCGGCTGCGCGACACCTTGTACCA 40005
QY 1109 GCATGACCGGTGCTGAGATCGAGCTGAAACGCGGTGACGACATCGGTGTTTCCGCC 1168
DB 40004 CGCGCGGACAAAGTTCGAGCTCACTCGGCGCAACGACAAACCGCTCTT----- 39951
QY 1169 CGATGGCAGCGTGCCTCGCGCTCGACGCGGCGAAATTTTCATGGCGCAGCATGTGGCGCTGAC 1228
DB 39950 CTTGAGGCGAAGAGATCTTCCACGCGGCGGAACTTTCACGCTCAGCGATCGCGTTCGC 39891
QY 1229 GTCGATGCTCGCTCGCCAGCGCGCTCACCGTTCTGCGCGGCTTTCGCGAGCGCGCATGTGC 1288
DB 39890 GATGAGACTCTGAGACCATCGCGCTCACCGTCTGCGGCTCTGCGCGAGCGGACATCAA 39831
QY 1289 AGTCTGACAGATGAAGGCTGAACCGGTGGGTGCGCGGCTTCTTCCACCGGCGGCGCGC 1348
DB 39830 CCGGCTCTGAAACCGGCACTCAGCTACGCGCTTCCCGGAGTTCCTCTGCTTCCGCGGAGCC 39771
QY 1349 CGGCTGAAATTCGCGCTTCATGCGGCGACAGTGCAGCGCGCGCTCTGCGCGGAGAT 1408
DB 39770 GGGCTGACAGCGATTCGCGGCGCGGCTGACCGCGCGCGGCTTCTGCGCGGAGGAA 39711
QY 1409 GCGAGCCACGCGGACCTGCTCGATCTCAATTCGATCTCAACGAGCGCGCAATCAGGATGT 1468
DB 39710 CCGGACGATCGCGCGCGGCGGACCGGAGCGCTCCCGTCCCAACGCGGACCAACGAGCGT 39651
QY 1469 GGTCTCGTTGGAGACATCGCGCGCGGCTCTGCGCGGAGAGATCGACCGTTGGCGGCGA 1528
DB 39650 GGTGAGCATGGGCGCTGATCTCGCGCGCGGCGGAGCGCGGCTCTGTGCGAACAACAA 39591
QY 1529 GATCTTGCAGTCTCGCTCTGCTTTCGACAGCTGCGGAGTGCCTGCGCGGCGGCGG 1588
DB 39590 GATCTCGCG-----GTGAGTACCTGCGCGCGCGGCGGCGGCTCGACATCTC 39537
QY 1589 CCTAGACGGGCTGTCTCCCGCGGGAAGAGCTGCTGAGCGGCTTCCGCGGAGCGAGTTC 1648
DB 39536 GTTCGAGCGCTTGAGCGCGGCGGAGGCGGAGCGGAGCGGCTGCGCGGCTGCTGCTCC 39477
QY 1649 GCGGCTTGAGAGCGACCGGCGGCTTGGGACAGGAAATTCGCGCGGCTTGC 1696
DB 39476 GACGCTGGGCTGCGACCGGTACATGGCGGCGGAGCATCGAGCTGCTGCGC 39429

RESULT 3
SCO939121      292100 bp  DNA  linear  BCT 16-APR-2005
LOCUS      Streptomyces coelicolor A3(2) complete genome; segment 18/29.
DEFINITION  AL939121 AL013137 AL160431 AL161691 AL161755 AL161803 AL353872
ACCESSION  AL356832 AL389898 AL450450 AL451182 AL589164 AL645882
VERSION  AL939121.1 GI:24429533
KEYWORDS  Streptomyces coelicolor A3(2)
SOURCE  Streptomyces coelicolor A3(2)
ORGANISM  Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomyces.
REFERENCE  1
```

AUTHORS Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C.H., Kieser, T., Lake, L., Murphree, L., Oliver, K., O'Neill, S., Rabinowitz, E., Rajandream, M.A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrell, B.G., Parkhill, J. and Hopwood, D.A.

TITLE Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2)

JOURNAL Nature 417 (6885), 141-147 (2002)

PUBMED 12000953

REFERENCE 2 (bases 1 to 292100)

AUTHORS Bentley S.D.

TITLE Direct Submission

JOURNAL Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces sequencing team, Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: sdb@sanger.ac.uk

COMMENT On or before Oct 30, 2002 this sequence version replaced gi:3449234, gi:7288050, gi:7320987, gi:7321265, gi:7649562, gi:8218130, gi:9367445, gi:11544744, gi:20520684.

FEATURES

source 1..292100

/organism="Streptomyces coelicolor A3(2)"

/mol_type="genomic DNA"

/strain="A3(2)"

/db_xref="taxon:100226"

97..103

111..3596

/gene="SCO4654"

/note="synonyms: rpoB, SCD82.26"

111..3596

/gene="SCO4654"

/note="SCD82.26, rpoB, DNA-directed RNA polymerase beta chain, len: 1161 aa; highly similar to SW:RPOC MYCTU (EMBL:L27989) Mycobacterium tuberculosis DNA-directed RNA polymerase beta chain (EC 2.7.7.6) RpoB, 1178aa: fasta scores: opt: 5849 z-score: 6361.5 E(): 0; 75.7% identity in 1169 aa overlap. Contains Pfam match to entry PF00562 RNA pol B, RNA polymerase beta subunit and match ot Prosite entry PS01166 RNA polymerases beta chain signature"

/codon_start=1

/transl_table=11

/product="DNA-directed RNA polymerase beta chain"

/protein_id="CAB77428.1"

/db_xref="GI:7248348"

/db_xref="GOA:Q9J0L0"

/db_xref="InterPro:IPR007120"

/db_xref="InterPro:IPR007121"

/db_xref="InterPro:IPR007641"

/db_xref="InterPro:IPR007642"

/db_xref="InterPro:IPR007643"

/db_xref="InterPro:IPR007645"

/db_xref="InterPro:IPR010243"

/db_xref="UniProt/Swiss-Prot:Q9J0L0"

/translation="MAASRNASTANNAASTAPLRISPAKIEPLEVFNLLAQTES FDMLLGNDKAWKVESAGQDVPVTKSGLEEIFEESPIEDFSGSNLFRDHRFEP PQNSIDBCRDFTYAFLFTVTAFTNNETGEIKSQTFVMDGDFPLMTKGTFFVINGTE RVVSVQLVSPGVVDFSDISDKTDKI FSAKII PSRGAWLEMEIDKRDVMGVRIIDKR KQSVTVLLKALGWTTEOILREFGVESMRATLEKHDTQGGDDALLDIYKLRGEPPT REAAQTLENIYKNPRYDLAKVGRYKVNKLGADELFDAGVLTDDVIATIKYLVKI HAGETVSGSGREIVVETDDIDHFGNRRIRNVGELLIQNVKRTGLAKMERVVRERTT QDVEAIPQTQLINIPRVASIKKEFFQSLSQFMDQNPLSGLTKRRNALPGGLS RERAFGRVDRMFCPIETPEGNGLIGLSASYGRINPFGFIETPYRKYVWG QVTDVYLTADBEDRVIAANAALGDDMRFAEVLVRRRGVEDVYVPGDDVMD VSPROMYSATAMPTLEHDANRALGMNMRQAVPLIKESPLCTGMGYESAADA GDVVKAKAGVQVQSADYLTWDDCTYLTTLAKESRSNQTQSVNQKVIYAEGRDI IEGVGLADGPATENGEMALGNLLVAFMPWEGHNYEDAIILSQLVQDDVLSIHIEE HEVDARDTKLGPEENLADLDDEGIIRIGAVVAGDILLVGKTPKGBT ELTPTEERLRAIRIEKAREVRDTSKYPHGEIYIGKVIQGVFRDEEGDELPGVYNQVLR VTYAQRKRKTGDGKLGRHGNKGVIISKINPIEDMPFLEDGTPVDIILNPLAVPSRMP

GOVLEIHLGLWLSASRGMDVSGLADEWAQRLOVIGADKVEPGTGNVATPVDFGAREDELALG LIQHTYIPNRDGERMVLPSGKARLFDRSGSEPRPEPISGVYMYILKLHLVDDKLHARS TGYSPMITQOPLGGKAPQGGQRGEMEWALEAYGAYALQELLTIKSDSDVTGRVKYV EALVKGENTPEGPISPEFKVLIKEMOSLCLNVEVLSDGMSIEMRDTDEDVFRAREEL GIDLSRRSESSVEEV"

411..3356

/gene="SCO4654"

/note="Pfam match to entry PF00562 RNA pol B, RNA polymerase beta subunit, score 1264.40, E-value 0"

2712..2750

/gene="SCO4654"

/note="PS01166 RNA polymerases beta chain signature"

3676..3682

3694..7593

/gene="SCO4655"

/note="synonyms: rpoC, SCD40A.01, SCD82.27"

3694..7593

/gene="SCO4655"

/note="SCD82.27, rpoC, DNA-directed RNA polymerase beta' chain (fragment), len: >1058 aa; highly similar to SW:RPOC MYCTU (EMBL:L27989) Mycobacterium tuberculosis DNA-directed RNA polymerase beta' chain (EC 2.7.7.6) RpoC, 1316aa: fasta scores: opt: 5139 z-score: 5536.5 E(): 0; 72.9% identity in 1067 aa overlap. Contains Pfam match to entry PF00623 RNA pol A, RNA polymerase alpha subunit. Contains also possible coiled-coil region approx. at residues 168..220

SCD40A.01, rpoC, DNA-directed RNA polymerase beta' chain (fragment), len: >279 aa; similar to C-terminal region of SW:RPOC BACSU (EMBL:L43593) Bacillus subtilis DNA-directed RNA polymerase beta' chain (EC 2.7.7.6) RpoC, 1199 aa; fasta scores: opt:801 z-score: 911.4 E(): 0; 53.0% identity in 247 aa overlap"

/codon_start=1

/transl_table=11

/product="DNA-directed RNA polymerase beta' chain (fragment)"

/protein_id="CAD55212.1"

/db_xref="GI:24429534"

/db_xref="GOA:Q8CJF1"

/db_xref="InterPro:IPR000722"

/db_xref="InterPro:IPR006592"

/db_xref="InterPro:IPR007066"

/db_xref="InterPro:IPR007080"

/db_xref="InterPro:IPR007081"

/db_xref="InterPro:IPR007083"

/db_xref="UniProt/Swiss-Prot:Q8CJF1"

/translation="MLDVNFDFELRIGLATADDIRQMSHGVEKKPETYINTYTLKPEKD GFCEKIFGTRDWCYCGKYKRVFRFGIICERCQVEVTRAKVRRMGMHIELAAPVT HIWFKGPSRLGYLLDLAPKDLKVIYFAAYMITFVDEERTRDLPSELAHSVVERQ QIBQRSDSLAEARAKLETDLAEAEAGKADVRKVRGAEAREMKOLRDRAQREIDR LDEVNRFKNLKVQDLLEGDELLYRELDRGTYPDCSMGAALQKRLSEFDLDEAER LREIITRGQKKTRALKRKLVSAFLQTSNPGMVLDCVPVIPDLRFMWQJDDGR FATSDNLDIRYRINRNRLKRLDIGAPEIIVNNEKRMLOEAVDALFDNGRGRPVT GPNRPLKSLMDLKGKQGRFRQNLGKRVYDSARVIVVGPQKLKCGCLKPMALIE LKFPFWMKRLVDLNLHQAQNKSAKRMVGRGTVVYDVLEEVIAHPHLLNRAPTLRGLG IOAFEPOLVEGKAIQIHPLVCTAFNADFDGDMAVHPLPSAEAAQAEARIIMLSSNNTL KPADGRVTMPQTDMLGLVFLTTDSEGRSPKGEAFGSSAEAIWAPADGLTLQAK IDIRFPVGTIPPRGFPPPAEEGPEWQGDFTLTKTLGRALFNEILLPDDVPDVEE VKGQSEIVNDLAEARYPKYIVAAITLNLKAAGFWATRSQVTAISDIIVPDAKKEI VKGYESQDQVQYERGLITKEERTQELIAITWKATNEVAEAMNDVPFVSMV NSGARMMOMROIAGMRGLVSNAKNETIIPRIKASFGREGLSVLEIVSTHGARKGLA DTLATFADSGYLTRRLVVSQDVIREEDCGTERGLKPIATRADGTLKAEADVETS VYARMLAEDVVIDGKVIAPANVLGDVLIDALVANGVEEVKTRSLITCESQVGTAMC VYGSLATGLKVIDGEAVGIIAASIGIEPGQLTNRFTHTGCGADDITQGLPRVVEHF EARTPKGVAFISEASGRVRIETEETKTKIWTPTDGSDETAFFISKARLLVGGHIV EVQKTLTGATNPDLVLRILGQAVOVHLVGEVQKVNYSQGSVHDKHIIITQMLR RVTIIESGDAELLPGELVERTKFTENRNVQSGHPASGRPOLMGITKASLASEML SAASFOETVLTDAINAKSDSLIGIKENVIIGKLIIPAGTGLSYRNIRVEPTEEAK AAMYSAVGDDIDYSPFGTSGQAVPLEDYDYGPNQ"

4621..6402

/gene="SCO4655"

/note="Pfam match to entry PF00623 RNA pol A, RNA polymerase beta subunit, score 1264.40, E-value 0"

misc_feature

misc_feature

RBS

gene

CDS

misc_feature

gene	polymerase alpha subunit, score 859.20, E-value 1.3e-254
	7703..8290
	/gene="SCO4656"
	/notes="synonym: SCD40A.02"
	7703..8290
	/gene="SCO4656"
	/note="SCD40A.02, possible integral membrane protein, len: 195 aa. Contains possible hydrophobic membrane spanning regions"
	/codon_start=1
	/transl_table=11
	/product="putative integral membrane protein"
	/protein_id="CAB81847.1"
	/db_xref="GI:7288052"
	/db_xref="UniProt/TREMBL:Q9L0K7"
	/translation="MSDDKPVNLLVSGVALLNLLILIMGCGSFAAFVLPVAGRAS NCGRPGVPCPGTGGVAGLAWAALALAHLMREGVGPMPAAAPVFLGAVLG ARPLSAVDFADSSRIIVMAVGVLLGLPGALYGRWLPGRWGAALLMGVRLDRGG PVPASPAQEGADRNILLALSILGAVLCAGGAALV"
gene	complement (8307..9611)
	/gene="SCO4657"
	/notes="synonym: SCD40A.03c"
	complement (8307..9611)
CDS	/gene="SCO4657"
	/note="SCD40A.03c, possible integral membrane protein, len: 434 aa; similar to TR:Q9EMX4 (EMBL:AL445403) Streptomyces coelicolor putative integral membrane protein SC1314.05, 396 aa; fasta scores: opt: 678 Z-score: 732.5 E1: 3.6e-33; 37.923% identity in 414 aa overlap. Contains possible hydrophobic membrane spanning regions"
	/codon_start=1
	/transl_table=11
	/product="putative integral membrane protein"
	/protein_id="CAB81848.1"
	/db_xref="GI:7288053"
	/db_xref="UniProt/TREMBL:Q9L0K6"
	/translation="WTAMDIAADTVTDGPGTERPAGGGAARSWRLPALPPGA RYELGYPALVRLDGLGVTSQSLTHAGDFGRNPRFGGAHWMDVLAITWDMGYLIO VHAHGYPALVRLDSDGLFTFQSSVAFPPYLPALIRMVSESTGLIGSGILSVVS KGRWATGIYAVTLLAGARAGTAAGLWAVAGVAVVSESLFVAIAAMTCYAVM SGVMTAGLITPAGLARSATVILGIAALVTLARPASRRHGVGPVYAMLVAP LGLLSYIAWGWSTGEATAYTLQREGWAHFFDYGATVDVLRNLAVRGDGYPFAPST PDLLSQLVLVALPFLIALMLRRRPLVLVATLTIIVTLGTOQMFGNTWRYLLPAPP
Query Match	13.1%; Score 242.4; DB 1; Length 292100;
Best Local Similarity	50.4%; Pred. No. 1.2e-22;
Matches	736; Conservative 0; Mismatches 676; Indels 48; Gaps 4;
Qy	215
Db	288054
Qy	275
Db	288114
Qy	335
Db	288174
Qy	395
Db	288234
Qy	455
Db	288294
Qy	515
Db	288354
Qy	575

Db 289449 ACCGACCGCCACCTGGCGC 289468

RESULT 4

AP006840 33/c

WPCOMMENT

Sequence split into 36 fragments LOCUS AP006840 Accession AP006840

Fragment Name Begin End

AP006840_00 1 110000

AP006840_01 100001 210000

AP006840_02 200001 310000

AP006840_03 300001 410000

AP006840_04 400001 510000

AP006840_05 500001 610000

AP006840_06 600001 710000

AP006840_07 700001 810000

AP006840_08 800001 910000

AP006840_09 900001 1010000

AP006840_10 1000001 1110000

AP006840_11 1100001 1210000

AP006840_12 1200001 1310000

AP006840_13 1300001 1410000

AP006840_14 1400001 1510000

AP006840_15 1500001 1610000

AP006840_16 1600001 1710000

AP006840_17 1700001 1810000

AP006840_18 1800001 1910000

AP006840_19 1900001 2010000

AP006840_20 2000001 2110000

AP006840_21 2100001 2210000

AP006840_22 2200001 2310000

AP006840_23 2300001 2410000

AP006840_24 2400001 2510000

AP006840_25 2500001 2610000

AP006840_26 2600001 2710000

AP006840_27 2700001 2810000

AP006840_28 2800001 2910000

AP006840_29 2900001 3010000

AP006840_30 3000001 3110000

AP006840_31 3100001 3210000

AP006840_32 3200001 3310000

AP006840_33 3300001 3410000

AP006840_34 3400001 3510000

AP006840_35 3500001 3566135

Continuation (34 of 36) of AP006840 from base 3300001 (AP006840 Symbiobacterium the:moph

Query Match 13.0%; Score 241; DB 1; Length 110000;

Best Local Similarity 51.8%; Pred. No. 2.5e-22;

Matches 786; Conservative 0; Mismatches 665; Indels 66; Gaps 8;

QY 183 GCGCTGAGCTGGATCGCCACATCGATCTGGACACGAGCCCATGCGCGTGGCGAGCGGCGC 242
Db 104276 GCGGTAGAACTCGGTGACATTTAACCTCCCGGAAGTCGTGGCTGTGCGCAGGACGCGC 104217
QY 243 GCGCGGATTTGCTTGTGCGCTCCGCGCGGACCGGTGCGCTGCGTCCGGAAGCGCGCTC 302
Db 104216 GCGCGCTGCTGCTGACCGCGGAGGTGGCGAGGCTAGCCAGAGCGAGATGGTG 104157
QY 303 GCGCGCTGCTATCCGCGAGGCGCGCATGTCTGCGAATGACAAACCGCGTTCGGTCCCTT 362
Db 104156 GAACGGCTGTGTCGGGAGCGCGCGCGGTCTACCGCATCACACCGGTCTTGGCAAGTTC 104097
QY 363 GCGAACCGCTGATCTCAGGTGAGAATGTCGGRACGCTGCGAGGCAATCTGTCCATCAT 422
Db 104096 AGCGACGTCCCATATCGGCAGAGACAGAGAATCTGACGCGCAATCTCTGATGAGC 104037
QY 423 CTGCGCAGCGCGTGGGACCGGTCTTGAATGAGACGCGCGCGCGCATGTTCTGGCG 482
Db 104036 CACGCTCGCGCTGCGAGAGCGCTCGCGCGGAAGTGGTGGCGGCGATGCTCTTCTG 103977
QY 483 CGTCTGCTGATCGCTCAGGAGAGCTTCGGGTGCCAGGAGGGAACCATCGCTCGCGCTG 542
Db 103976 GCGCCCGAGGCGCTCTCCCGGGGCCACTCGGGCATCCGGGCGGAGACCTCTGGAGATGCTC 103917

QY 543 ATCGACCTGCTCAATTCGAGCTCGTCCGCGCGGTTCCAGCCGCGGACGGTGGGGCGC 602
Db 103916 GTGGGTTCTCAATCTCGGCTCACCCCGTGGTGGCGAGCAGGCTCCCTGGGCGCC 103857
QY 603 TCGGGTGACCTGACACCGCTTGGCGCATATGGTGTCTGCTCCAGGCGCGGGGAGATTC 662
Db 103856 AGCGGCGACCTCGCGCGCTCGCCACATGAGCCTTCGCTGATCGGGTTGGGTGAGGCC 103797
QY 663 CTGACCGGACCGGAGCGGCTTGAAGGCGGAGAGGGTCTCGGCGGAGCGGCTGCAA 722
Db 103796 GTTGTCAATGGCGA---GCGGCTGTCGCGCGCGGAGCGCTGCGAGCGGTCGCGCTCCGG 103740
QY 723 CCGCTCGATCTCTCCCATCGCGATGCACTGCGCTGTGTCAACGCGAGCCTCGCCCATGACC 782
Db 103739 CCGCTGAGCTGACGGCCAAAGGAGGCTGGCTGCTGATCAACGCGCACCCAGGCGATGACC 103680
QY 783 GGGATCGCGCTGCTGTAATGCTCAAGCTGCGCGCATCTCGGCAACTGGGCGGTGGCGTTG 842
Db 103679 GCGCTGGGTCCTCGGCTCCAGATGCTCAGTGTCTGCTGAAACCGCGGACATCGCG 103620
QY 843 ACGCCCTGCTTGGGAAATGCTGAGAGGCGGACCGAGGATGGGCGCGGCACTGTCC 902
Db 103619 GCGGCCATGACCGGCGGAGCGCTGGGCGCGATCCCGCGGCTTGGGACCGCGGCTGCGAG 103560
QY 903 GACTGCGCGCGCATCCCGGACAGAGGACGCGCAGCAGAGGCTGCGCGCCCGCGTGGAC 962
Db 103559 GCGCTGCGGCTCCACACGGGCGAGCGGCTGCGAGAACCTGCGCGCGC----- 103508
QY 963 GGCAGCGCGCGGTGGTTCGCGCACGCTATTGCCGAGCGGAGGCTCGACGCGCGGATATC 1022
Db 103507 -----TGACCGAGGGTTCGCGCTCAC-----ACC 103482
QY 1023 GGCAGCGGACCGGAGGGGGGAGGATGCTCTACAGCTGCGCTGCGCTCCGAGGTTCTC 1082
Db 103481 CGACCGGGGAGATGCGCACCCAGGATCCGTCACCTGCGCTGCTTCGCGAGGTGCAC 103422
QY 1083 GCGCGCGGCTTCGACACGCTCGCATGGCATGACCGGCTGTGACGATCGAGCTGAACGCG 1142
Db 103421 GCGCCAGCGGACCGGATCGAGCAGTGGCGAGGCTGTGACTGGAGATGAAGCC 103362
QY 1143 GTGACCGCAATCCGGTGTTCGCGCGGATGGCAGCGTGCCTCCCTGCAAGGGGCAAT 1202
Db 103361 GTACCGCAATCCCTGCTCTTCGCGGACGACGACGAG---GTGATCTCCGCGGCAAC 103305
QY 1203 TTGATGGCCAGCATGTGGCGCTGAGCTCGATGCGCTCGCCACGCGCTCACGTTCTG 1262
Db 103304 TTCCAGGTCAGCGGTCGCGCTGGCGCTGAGTACTTGGCCATCGCCGTCGCGGAGCTC 103245
QY 1263 GCGGCGCTTCGCGAGCGCCAGATTGACGCTTCACAGATGAAGGCTGAACCGTGGGCTG 1322
Db 103244 GCGCATATCGCCAGCGGCGCATCGAGCGACTGGTCAACCCACGATTTTCG---GGGCTG 103188
QY 1323 CCCCCCTTCTCCACCGGGGCCCCGCGGTTGAAATTCGCGCTTCATGCGGCGACAGGTG 1382
Db 103187 CCGGCTTCTCACCC---GCAACGCGGGGTGATTCGCGGCTGATGATCAACCCAGTAC 103131
QY 1383 ACGGACCGCGCTCTTGGCCGAG---ATCGGAGCCACGCGGACTTGGCTCGATCCATTCG 1439
Db 103130 ACGGCGCGCTGCTGTGTGAGCGAGAACAGGTGCTCGCTCACCCGCGCAGCGTGGACTCC 103071
QY 1440 ATCTCCACAAACCGCGCAATCAGGATGTGGTCTCGCTTGGGACCATCGCGCGCGGCTC 1499
Db 103070 ATTCTGTCAGCGCAACAGAGGAGCACGCTGTGATGGGACACCGGCGCGCCGCAAG 103011
QY 1500 TGGCGCGAGAGATCGACCGTTGGGCGGAGATCTTTCGATCTCGCTCTCTGTCTTGA 1559
Db 103010 GCGCGGAGGTGATGCGCAACGTCGCGCGGTTGCTGSCCATCGAGCTGTGTGCGCGCC 102951
QY 1560 CAAGCTCGGAGCTGCGCTCGGCGAGGGGCTAGACGGGGTGTCTCCCGCGGGGAGAG 1619
Db 102950 CAGGCTCTGAGATTCTGGGCGCGGAGCGGCTCGCCCGCGCCACGCGCGCGCTACGC- 102892

Qy 1620 CTGGTGCAGGCCCTGCGGAGCAGTTCCCGCGCTTGAGACGGACCGGCCCTCTGGACAG 1679
Db 102891 -----GGCGATCCGGGAGCGGTGGCGCGCTGAGCGGCGACCGGGTGTCTGGCGCG 102840
Qy 1680 GAAATTCGCGGCTTGC 1696
Db 102839 GACATTGAGGCCCTGCG 102823

RESULT 5

AP006840_34/c

WPCOMMENT

Sequence split into 36 fragments LOCUS AP006840 Accession AP006840

Fragment Name	Begin	End
AP006840_00	1	110000
AP006840_01	100001	210000
AP006840_02	200001	310000
AP006840_03	300001	410000
AP006840_04	400001	510000
AP006840_05	500001	610000
AP006840_06	600001	710000
AP006840_07	700001	810000
AP006840_08	800001	910000
AP006840_09	900001	1010000
AP006840_10	1000001	1110000
AP006840_11	1100001	1210000
AP006840_12	1200001	1310000
AP006840_13	1300001	1410000
AP006840_14	1400001	1510000
AP006840_15	1500001	1610000
AP006840_16	1600001	1710000
AP006840_17	1700001	1810000
AP006840_18	1800001	1910000
AP006840_19	1900001	2010000
AP006840_20	2000001	2110000
AP006840_21	2100001	2210000
AP006840_22	2200001	2310000
AP006840_23	2300001	2410000
AP006840_24	2400001	2510000
AP006840_25	2500001	2610000
AP006840_26	2600001	2710000
AP006840_27	2700001	2810000
AP006840_28	2800001	2910000
AP006840_29	2900001	3010000
AP006840_30	3000001	3110000
AP006840_31	3100001	3210000
AP006840_32	3200001	3310000
AP006840_33	3300001	3410000
AP006840_34	3400001	3510000
AP006840_35	3500001	3566135

Continuation (35 of 36) of AP006840 from base 3400001 (AP006840 Symbiobacterium the moph

Query Match 13.0%; Score 241; DB 1; Length 110000;
Best Local Similarity 51.8%; Pred. No. 2.5e-22;
Matches 786; Conservative 0; Mismatches 665; Indels 66; Gaps 8;

Qy 183 GCCGTGAGCTGTGATCGCCATCGATCTGGACGAGGCCCATCGCTGGCGAGCGGGCGG 242
Db 4276 GCGGTAGAACTCGGTGACATTTAACTCCCGGAGTGTGGTGTGCTGCCAGGACGGC 4217
Qy 243 GCGCGGATGTCTTGGCCCTCCCGCGCGGACCGGTGCGGTGCGTGGAGACGCGGCTC 302
Db 4216 GCGCGCGTGTGTGACGCGCGGAGTGGCGGCGGTAGCCAGAGCGAGATGGTG 4157
Qy 303 GGCCTGTATCCGAGGCGCGCATGTCTACGACTGACAAACGGCTTCGGTCCCTT 362
Db 4156 GAACGGCTGTGGGAGCGCGCGGTCTACGCGCATCACACCGGTTTGGCAAGTTC 4097
Qy 363 GCGAACCGCTGTATCTCAGGTGAGAAATGTCCGAAACGCTGACAGGCAATCTTGTCCATCAT 422
Db 4096 AGCGACGTCCCATATCGGACAGACAGACAGACATGACAGCGCATCTCTGATGAGC 4037
Qy 423 CTGGCCAGCGCGGTGGGACCGGTGTGACTGGAGCGAGCGCGCGGCATGGTCTCGCG 482

ASLAEFWAHLIGGEVAVOTPGCAVEVGSLLHLAGVRVSGYRPTWPSDRSQIHLDL
TVDDLDTAEHEAVRLGATREHQPSPERTVRLDRDPAGHPFLRA"

complement(11363..12862)
/note="similar to Streptomyces coelicolor putative
rhamnose kinase in NP_625114"

/codon_start=1
/transl_table=11

/product="probable sugar kinase"
/protein_id="AAZ23051.1"

/db_xref="GI:71068196"

/translation="MTTPPSALASTFAAVDLGATSGRVIGRVGRNLDLTAHRF
PNTVRLPDGLRMDVLFQGLDGLREARSGVAVSGVDTWAVDHGLLDADGALLG
HPFYRDSKTDAARVHSLVPELRYITGLHLPFTTVFQLAAEGTGVOKAARTL
LLIPDLLJHMLTGSVASETNASTTGLFDARTGWSQDLRLSLDPLGQLPQLRPGD
PAGTLLPVAAYTGLGPRIPVTIVASHOTASAVPAIEPFDFAISCTGWSLAGELE
DAPVLTESRAANFTNERGIDGTIRYLNRIMGMLLEECRTWRNENLPAHLSALLAD
AARAEFASVDPDPAPVAFQDMPARIRDYCARTRVPGSQGAIVRVLESLLAH
RRALRQAALAGRLDRIHLVGGSRNDLLCOLTADATGLPVAGPTATAGNLILVQ
ARATGLVDLADMRLELIASHTOHLRHYTPRGDGTAWHAAARLDLPAGHR"

complement(13059..15098)

/note="similar to Streptomyces avermitilis putative
sorbitol-6-phosphate 2-dehydrogenase in NP_828590"

/codon_start=1
/transl_table=11

/product="probable dehydrogenase"
/protein_id="AAZ23052.1"

/db_xref="GI:71068197"

/translation="MATHPEVAALIGRAHLRSGDPRNTNYAGGNASAKGIIVTDPVTGS
NTELMWVKGSGDGLTITDEGLVALRLDLRALQDVPVGEREDEMVAAPDCHLHGRG
GAAPSDTAMHGLVDVDPHVDPSGIALACAADGEKLTAEVCYGAKEVAVPWRPGR
OLGLDIAAEKHAEPVGVGLGHGIGTAWGSAQCEANALWMIRTAELFVREGRE
PFGPLEGALGAHERERAAALAPVIRSLASQDRPOVGHFTSEVVLDFLAHRRP
HLAALGTCSPDHFIRKVRPMVLDLPADAPIDEAVARLKEIHAEYREYAAAYVRRHAT
AASPMRGADPAIIVLPVGMFSGRDKQRTARVAGEFYVNAIVMRGAEAVSSYAPIE
BAEKFRIEYWELEAKLRMPKPKPLATRVALTGAGSIGKAVARLAAEGACVVA
DLNGNAAVEALEBGLDQAVAVTVDTSBQIIAAAPREAVLAFGVLDLVNNAISV
SKPLETTARDMDLQHDIMARGSEFLVSEALVMTVOGLGDIIVYIVSKNAVFAGPN
IAYSATKDAQHVRLLAAELGEHGVIRVGNVPGVVRGSGIPAGWGQAQRAAYTGD
BEKIGEFYARTLKLREVLPEHVAANAVFALTGELSHSTTGLHVPVDAGVAAAFYLR
complement(15190..16359)

/note="similar to Streptomyces avermitilis putative sugar
isomerase in NP_828591"

/codon_start=1
/transl_table=11

/product="probable sugar isomerase"
/protein_id="AAZ23053.1"

/db_xref="GI:71068198"

/translation="MLDVRDAKALKTQTVETPSWGVNSGTRFKFPAOPGVPTPW
EKLEDAKVEHFTGVTTPKYSLSHIPDRVDDYTALAAAYAAERGLRLGAINSTFODDY
RLGVSCHPDARRRAGVHLLUECDVIMDATGSTDLKLFADGTNYPOGDDVVARQDRL
AALAEVYERLGEERLLELYKJFEPAPFTTDPDWTGTSYAHCLKLGPKAQVVDYTG
HAPGTSEFIVAFLLREGLGDFNSRFYADDDLMVGSSDPFQFPRIMHEVTVNGGL
APAANVPMLOQCHNIEAKIPAVIRSVNMVQEAETAKALLVDLDAALAAQREGDVLAAN
AALMDAVNTDVRPILLREBQGLAPELAAYASGWRERATATERVGEQAGWCA"
16364..18278

/note="similar to probably glycosidase; reading frame is
interrupted; similar to Streptomyces narbonensis NbmF
encoded by GenBank Accession Number AAM88355"

18507..19304

/note="similar to Streptomyces avermitilis hypothetical
protein in NP_822661"

/codon_start=1
/transl_table=11

/product="hypothetical protein"
/protein_id="AAZ23054.1"

/db_xref="GI:71068199"

/translation="MRRGRCAIYASARGKPRHSLTPNISLNLGLSVITIGISPDVA
DAFDALIAVTATGLRGHEPWPATAALTDLQRLAAGTWQPADESIDPRLEAHTAY
RSFGTNRIRIPSDVLAGRMVKKGTLPINPAVDVSYNAVSIRHGLPAGAPDLPHVTG
DVDIRYADGSTFTPLAGEPDTVENPKPEVYADSDVLTRHWNHRDHRTRVTEGST
HVAFVLETLHATRDGHLKTAALVELQELLAPHAEQTTVHYLAPEAGA"

12.8%; Score 236.4; DB 1; Length 126933;

Query Match

Best Local Similarity 51.0%; Pred. No. 9.3e-22;
Matches 697; Conservative 0; Mismatches 626; Indels 45; Gaps 4;

Qy 213 GACCAGGCCATATGCGTGGCGAGCGCGCGCGGATTTGCTTTCGCCCTCCGCGCGGC 272

Db 4864 GACGAAGTCGTGTCGTCGCGCGCTCTCGGGACCCCGGTGGTCTCTGCTCCCGAAGCGCTG 4805

Qy 273 GACCGTGCCTGCTGCTCGAAGCGCGGCTCGGCGTGTATCTCGGAGCGCGGCATGTC 332

Db 4804 AAGGAGATCGCCCGCAGCCGGAGGTCTACAGGCTCTCGCCGACGACGCCAAGCCGAC 4745

Qy 333 TACGAGCTGACAAACCGGCTTCGGTCCCTTCGGAACCCCTGATCTCAGGTGAGAATGTC 392

Db 4744 TACGGCTCTCACCGGTTTCGGTGGCTGGCCACCCGCCACATCCCAACGAGGTGCGC 4685

Qy 393 CGAAGCTGACGGCAATCTTTGTCCATCATCTGGCCAGCGGCTGGAGCCGGTGTTCGAC 452

Db 4684 GCCAGCTCCAGCGCGTCTGCTGCGCTCGCACCGCGGGCTCGGGACCCGAGGTGCGAG 4625

Qy 453 TGGAGAGCGCGCGGCATCGTTCGCGGCTGTGCTGATCGATCGACCTGCTCAATTCGAGCTCGCTCCG 512

Db 4624 CGCAGGTGCTCGCGCGCTCATGTCTGCGACTGTCCACGCTCGCCACCGCGCGGTACC 4565

Qy 513 GGTGCCAGGAGGAGCATTGCTCGCTGATCGACCTGCTCAATTCGAGCTCGCTCCG 572

Db 4564 GGTGTCCGACCGAGAGCGCCGCGGTACCGCGCTGCTCAACCGCGGATCACCCCG 4505

Qy 573 GCCGTTCACCGCGCGCACCGTGGCGCGCTGGGTGACCTGACACCCCTTTCGCGCATATG 632

Db 4504 GTCTGTCACGAGTACGGCTGCTCGTGGTTCGCGTACCTCGCCCTTCGCGACTGCG 4445

Qy 633 GTGCTTGCCTTCAGGGCGCGGAGAGATTCTTGGACCGGGACGGGACCGCGGCTTTCGACGCG 692

Db 4444 CGCTGCGCGCTCATGGGCGAGGAGACGCTCGCGCACCGCGCGCGCTGCGGAGCGCGCG 4385

Qy 693 GCAGAGGGCTCGCGCGCGGAGCGCTGCAACCGCTGCAATCTCTCCATCGCGATGCACTG 752

Db 4384 GCGAGGCGCTCGCGGAGCGGGGATCACCCCGTCTGCTTGGAGGAGAAAGAGGGCGCTG 4325

Qy 753 GCGTGTGTAACGGGACCTCGCCATGACCGGGATCGCGCTGGTGAATGCTCACCGCTCG 812

Db 4324 GCCCTCATACCGGACCGAGCGCATGCTCGGCGTGTGGCGCTGGCGCCCGACGATCTG 4265

Qy 813 CGCATCTTGGCAACTGGCGGCTGGCGTGTGACGGCCCTGCTTGGGGAATGTCTGAGAGCG 872

Db 4264 CGGCGCTCTCTCCGACCGCGCGACATCGCGCGCGCATGACGCTGAGCGCTGAGCGCGCAGCTCGCGC 4205

Qy 873 CGGACCGGACATGGCGCGCGGACATGTCGACCTCGCGCGCATCCCGGACAGAGGAC 932

Db 4204 ACCGACCGGCTTTCGCGCGCGACCTTCAGGCGCTTGGCGCCCGCACCGGGCCAGCGCCGAC 4145

Qy 933 GCCCGAGCGAGGCTCGCGCGCGCGTGGAGCGGACCGCGGGTGGTCCGCGCATGCTCAT 992

Db 4144 AGCGCGCCCAATCTTGGCGCGCTGCTCGCGGCTCGCGGATCGTGGCCAGCCACAGGGC 4085

Qy 993 GCCGAGCGGAGGCTGACGCGCGCGATATCGGAGCGGAGCGGGGCGGAGCGGATGCC 1052

Db 4084 CCGGAGTGCACGCGCGT-----CGAGGACGCC 4058

Qy 1053 TACAGCTGCGCTGCGCTCGCAGGTTCTCGGGGCGGCTTCGACACGCTCGCATGGCAT 1112

Db 4057 TACTCGCTCGCTGCAACCGCCAGGTCCAGGCGCCCGCGTGCACCCCTCTCCACGCG 3998

Qy 1113 GACCGGTGTGTGACGATCGAGCTGAACCGGCTGACCGCAATCCGGTGTGTTTCGCGCCGAT 1172

Db 3997 GAGCTGTGCGCGCGCGGAGCTGCGCTTCGCCATCGACAAACCCCGTCTCACGCGTAC 3938

Qy 1173 GGCAGCTGCGCGCTTCGACGCGGGGCAATTTTATGCGCCAGCATGTGGCGCTGACGCTCC 1232

Db 3937 GCGCGGCTG-----GAGAGCAACGCGCACTTCCACGCGCTCGCGGTGCGCGCGCTGCTC 3884

Qy 1233 GATCGGCTGCCACGCGCGCTCACCGTTCGCGGGCGCTTCGCGGAGCGCCAGATTGCACTG 1292

Db 3883 GACTTCCTCGCCATCTCGTCCGCGAGCTCGCTCGATCTCCGAGCGCGCAGCAGCCG 3824
Qy 1293 CTGACAGATGAAGGCTGAACCGTGGCTGCCCCCTTCTCCACCGGGGCCCGCCGGG 1352
Db 3823 TTCCTGACCGCGCGCAATCGCGGCTCAACGCTTCTTCGGC---GAGCACCCCGGA 3767
Qy 1353 TTGAATTCGGCTTCATGGGCGCACAGGTGACGGCGACCGGCTCTCTGGCCGA---GATG 1409
Db 3766 GTGGACTCCGSCACATGATCGCGCAGTACACCCAGCGCGGATCGTCTCGAGCTGAAG 3707
Qy 1410 CGAGCCAGGAGACCTGCTCGATCCATTCGATCTCCAGAAAGCGCGCAATCAGGATGTG 1469
Db 3706 CGGCTCGCGTGGCGGCTCCACGACTCGATCCCTTCCAGCGCCATCGCAGGAGCAC 3647
Qy 1470 GTCTCGCTTGGACCATCGCGCGGCTCTGCGCGAGAGATCGACCGTTGGCGCGAG 1529
Db 3646 GTCTCCATGGCTGGTCCGCGCGCCGCAAGTGGCGCGCCCTCGACGGTCTCGGCGG 3587
Qy 1530 ATCCTTGGATCTCTGCTCTGTCTTGTGCAAAAGCTGGGAGCTGCGC 1577
Db 3586 GTCGTGCGGTGGAACCTCTCACCGCGCGCGGCTCGACCTGCGC 3539

RESULT 7

BA000030_41/c

WPCOMMENT

Sequence split into 91 fragments LOCUS BA000030 Accession BA000030

Fragment Name	Begin	End
BA000030_00	1	110000
BA000030_01	100001	210000
BA000030_02	200001	310000
BA000030_03	300001	410000
BA000030_04	400001	510000
BA000030_05	500001	610000
BA000030_06	600001	710000
BA000030_07	700001	810000
BA000030_08	800001	910000
BA000030_09	900001	1010000
BA000030_10	1000001	1110000
BA000030_11	1100001	1210000
BA000030_12	1200001	1310000
BA000030_13	1300001	1410000
BA000030_14	1400001	1510000
BA000030_15	1500001	1610000
BA000030_16	1600001	1710000
BA000030_17	1700001	1810000
BA000030_18	1800001	1910000
BA000030_19	1900001	2010000
BA000030_20	2000001	2110000
BA000030_21	2100001	2210000
BA000030_22	2200001	2310000
BA000030_23	2300001	2410000
BA000030_24	2400001	2510000
BA000030_25	2500001	2610000
BA000030_26	2600001	2710000
BA000030_27	2700001	2810000
BA000030_28	2800001	2910000
BA000030_29	2900001	3010000
BA000030_30	3000001	3110000
BA000030_31	3100001	3210000
BA000030_32	3200001	3310000
BA000030_33	3300001	3410000
BA000030_34	3400001	3510000
BA000030_35	3500001	3610000
BA000030_36	3600001	3710000
BA000030_37	3700001	3810000
BA000030_38	3800001	3910000
BA000030_39	3900001	4010000
BA000030_40	4000001	4110000
BA000030_41	4100001	4210000
BA000030_42	4200001	4310000
BA000030_43	4300001	4410000
BA000030_44	4400001	4510000

BA000030_45	4500001	4610000
BA000030_46	4600001	4710000
BA000030_47	4700001	4810000
BA000030_48	4800001	4910000
BA000030_49	4900001	5010000
BA000030_50	5000001	5110000
BA000030_51	5100001	5210000
BA000030_52	5200001	5310000
BA000030_53	5300001	5410000
BA000030_54	5400001	5510000
BA000030_55	5500001	5610000
BA000030_56	5600001	5710000
BA000030_57	5700001	5810000
BA000030_58	5800001	5910000
BA000030_59	5900001	6010000
BA000030_60	6000001	6110000
BA000030_61	6100001	6210000
BA000030_62	6200001	6310000
BA000030_63	6300001	6410000
BA000030_64	6400001	6510000
BA000030_65	6500001	6610000
BA000030_66	6600001	6710000
BA000030_67	6700001	6810000
BA000030_68	6800001	6910000
BA000030_69	6900001	7010000
BA000030_70	7000001	7110000
BA000030_71	7100001	7210000
BA000030_72	7200001	7310000
BA000030_73	7300001	7410000
BA000030_74	7400001	7510000
BA000030_75	7500001	7610000
BA000030_76	7600001	7710000
BA000030_77	7700001	7810000
BA000030_78	7800001	7910000
BA000030_79	7900001	8010000
BA000030_80	8000001	8110000
BA000030_81	8100001	8210000
BA000030_82	8200001	8310000
BA000030_83	8300001	8410000
BA000030_84	8400001	8510000
BA000030_85	8500001	8610000
BA000030_86	8600001	8710000
BA000030_87	8700001	8810000
BA000030_88	8800001	8910000
BA000030_89	8900001	9010000
BA000030_90	9000001	9025608

Continuation (42 of 91) of BA000030 from base 4100001 (BA000030 Streptomyces avermitilis)

Query Match	12.6%	Score 232.8;	DB 1;	Length 110000;
Best Local Similarity	50.0%;	Pred. No. 2.8e-21;		
Matches 730;	Conservative	0;	Mismatches 682;	Indels 48; Gaps 4;

Qy	215	CGAGGCCATGCGTGGCGGAGCGCGCGGATGTCCTTGGCCCTCGGGCGCGG 274
Db	39677	CGAGCTTCTCGCGTGGCGGCGCGGATCGAGCTCTCGGGGAGCGGCTGAC 39618
Qy	275	CCGGTCCGTCGTCGGAAGCGGCTCGCGGTGTCATCCGCGAGCGCGCCATGCTA 334
Db	39617	CGCCCTCGCGCGCGCGGATCGTGGACGCGTGGCGGCAAGCCGAGCCGTGTA 39558
Qy	335	CGGACTGACAAACCGGCTTCGGTCCCTTGGAAACCGCTGATCTCAGGTGAGATGTCG 394
Db	39557	CGGGTGAGCACCGGCTTCGGGGCCCTGGGACCCCGGCACATCAGCAGAGCTGCGGGC 39498
Qy	395	AACGTCGAGGCCAATCTTGTCATCATCTGGGCGAGCGGCGTGGGACCGGTGTTGACTG 454
Db	39497	GCAGCTCCAGCGCAACATCGTCCGCTCGCACGCGCGCGGATGGGTCCGAGGGTGGAGC 39438
Qy	455	GACGCGGCGCGCCATGTTCTGGCGGTCTGCTGTCGATCGCTCAGGAGCTCCGG 514
Db	39437	GGAGTGTACGGGCGCTGATGTTCTTGCAGCTGAACACGCTCTGTCGGGCGCACACGG 39378
Qy	515	TGCCAGCGAGGGACCATCGCTCGCCTGATCGACCTGCTCAATTCCGAGCTCGTCCGGC 574

Qy 529 CCATCGCTCGCTGATGACCTGTCTCAATTCCGAGCTCGCTCCGCGCGTTCCTCCAGCGCG 588
Db 71481 TGATGGAGCGCTCGTACGCTGTTCAACGGCGAGTGTCTGCCCTCATTTCCGTCGAAG 71540
Qy 589 GCACGGTGGCGCGTGGGTGACCTGACACCGCTTGGGCATPATGTTGCTCTGCCCTCCAGG 648
Db 71541 GCTCGTGGCGCGTGGCGGACCTTGGCGGCTCGCGCATGTGCGCGGTGCTGCTCG 71600
Qy 649 GCCGGGAGACTTCTTGACCGGAGCGGAGCGGCTTGACGGCGCAGAGGGCTCCGGC 708
Db 71601 GCATCGCGACGTGTTCATCCGCGCGA---GCGCGGAGCGCGCGCAAGGGTGGGTG 71657
Qy 709 GCGGACGCTCAACCGCTGATCTCTCCATCGCATGCACTGCGCTGTGTCAACGGGA 768
Db 71658 TCGCGGGCTCGCGCGCTCAGCTCGAAGCAAGAGGGCTCGCGTGTGAACGGCA 71717
Qy 769 CCTCCGCCATACCGGGATCGCGTGTGTGAATGTCTACGCTCGCGCATCTCGGCAACT 828
Db 71718 GCGAGGCTCGACCGCTCGCGCTCGACAACTGTTTCGGATCGAGGACCTGTACCGCA 71777
Qy 829 GGGCGGTGGCTTGAACCGCTGCTTGGCGAATGTCTGAGAGGCGGACCGAGGCATGGG 888
Db 71778 CGGCGCTCGTTCGGGCGCGCTGTCTGCTCGACGGCGGCGGCTCGGTGAAGCCGTTTCG 71837
Qy 889 CCGCGGCACTGTCGACCTCGCGCGCATCCCGACAGAGGAGCGCGCAGCGAGGCTGC 948
Db 71838 ACGCGCGCATCCAGAGCTGCGCGGCATCGCGGCAGATCGACGCGCGCGCGCTGAC 71897
Qy 949 GCGCCCGCGTGGACGCGAGCGCGGGTGTGTCGCGCACGTCAATTGCGGAGCGGAGCTCG 1008
Db 71898 GGTGCTGCTCGACGGCTCGGCGATCAACGTGTGCGACCGC----- 71938
Qy 1009 ACGCGGGGATATCGGGACGAGCGCGGCGGGCGAGGATGCTCTACAGCTGCGCTGCG 1068
Db 71939 -----GATTGCGACAAGGTGCAAGGACCGGTACAGCCTGCGCTGCC 71978
Qy 1069 CTCGCGAGGTTCTCGGCGCGGCTTCGACACGCTCGCATGSCATGACGGTGTGACGA 1128
Db 71979 AGCGCGAGTGATGGCGCGTGTCTGACACAGATCCGCGCAGCGCGCGCGTGTGCTCA 72038
Qy 1129 TCGAGCTGAACCGGTGACGACCAATCCGCTGTTTTCGCGCCGATGGCAGCGTGCCTCC 1188
Db 72039 TCGAGGCGAACCGGTGTCGACAAACCGCTGATCTTCCCGGACACGCGCGAG---GTGC 72095
Qy 1189 TGCACGGGGCAATTTGATGGGCGAGCATGTGGCGCTGAAGTCGATGCGCTCGCCACGG 1248
Db 72096 TGTCCGGCGGCAATTTCCACGCGGAGCCGCTCGCGTTTCGCGCGCGCAATCTCGCGATCG 72155
Qy 1249 CCGTCACGTTCTCGCGGCTTCGCGAGCGCCAGATTGCAAGTTCACAGATGAAAGGC 1308
Db 72156 CCGCGCGAGATCGGCGCTTCGCGAGCGCCGATCGCGCTTTGATGACGCGAGCG 72215
Qy 1309 TGAACCGTGGCTGCCCGCTTCTTCCACCGGGGCGCGCGGGTTGAATTCGCGCTTCA 1368
Db 72216 TCTCGGCGCTCGCGC-----TTTCTCTGTGAAGGAGCGGCGGTGAACCTCGGCTTCA 72269
Qy 1369 TGGGCGCAGGTGACGGCGACCGCGCTCTTGGCGGAGATCGGAGC---CAGCGGACTGT 1425
Db 72270 TGATCGGCGACGTGACGCGCGCGCTCGCGTCCGAAACCAAGACGCTCGCGATCCGG 72329
Qy 1426 CTTGATCCATTTGATCTCCACGAACCGCGCAATCAGGATGTGTCTCGTTTCGGACCA 1485
Db 72330 CGTGGTGGATTCGTCGCGAGCTCGGCGAACAGGAGACCACTGTGTGATGCGGAGCT 72389
Qy 1486 TCGCGCGCGCTCTTGGCGGAGAGATCGACCGTTGGGCGGAGATCCTTTCGATCCTCG 1545
Db 72390 TCGCGCGCGCAAGCTACGAGCATCGCGGAGACGTTCGCGAAACATCTTCGCGATCGAGC 72449
Qy 1546 CTCTCTGTGTCACAGCTCGGAGCTGCGCTCGGCGAGCGGCTAGACGGGGTGTCTC 1605
Db 72450 TGCTCGCGCGCGCAAGGCGTGCACCTGC-----GCGCGCGCGACGCAACGAGCC 72500

Qy 1606 CCGCGGGAAGAAAGCTGCTGAGGCCCTGCGCGAGCAGTTCCGCGCGCTTGAGACGAGCC 1665
Db 72501 CGGCGCTGAGCAGCGCATGAAGACGATTCCGCGGAGCTCGCGCATCAGATCTCGACC 72560
Qy 1666 GGCCCTTGGGACAGGAAATTGCCCGCTTCTAGCACCT 1705
Db 72561 ACTACTTCGCGCGCGACATCGCGTGGTTCGCGGCGCGT 72600

RESULT 9
BX571965_28/c
WPCOMMENT
Sequence split into 41 fragments LOCUS BX571965 Accession BX571965

Fragment Name	Begin	End
BX571965_00	1	110000
BX571965_01	100001	210000
BX571965_02	200001	310000
BX571965_03	300001	410000
BX571965_04	400001	510000
BX571965_05	500001	610000
BX571965_06	600001	710000
BX571965_07	700001	810000
BX571965_08	800001	910000
BX571965_09	900001	1010000
BX571965_10	1000001	1110000
BX571965_11	1100001	1210000
BX571965_12	1200001	1310000
BX571965_13	1300001	1410000
BX571965_14	1400001	1510000
BX571965_15	1500001	1610000
BX571965_16	1600001	1710000
BX571965_17	1700001	1810000
BX571965_18	1800001	1910000
BX571965_19	1900001	2010000
BX571965_20	2000001	2110000
BX571965_21	2100001	2210000
BX571965_22	2200001	2310000
BX571965_23	2300001	2410000
BX571965_24	2400001	2510000
BX571965_25	2500001	2610000
BX571965_26	2600001	2710000
BX571965_27	2700001	2810000
BX571965_28	2800001	2910000
BX571965_29	2900001	3010000
BX571965_30	3000001	3110000
BX571965_31	3100001	3210000
BX571965_32	3200001	3310000
BX571965_33	3300001	3410000
BX571965_34	3400001	3510000
BX571965_35	3500001	3610000
BX571965_36	3600001	3710000
BX571965_37	3700001	3810000
BX571965_38	3800001	3910000
BX571965_39	3900001	4010000
BX571965_40	4000001	4074542

Continuation (29 of 41) of BX571965 from base 2800001 (BX571965 Burkholderia pseudomallei)

Query Match 12.4%; Score 230; DB 1; Length 110000;
Best Local Similarity 50.8%; Pred. No. 6.4e-21;
Matches 752; Conservative 0; Mismatches 665; Indels 63; Gaps 6;

Qy 229 TGGCGAGCGCGCGCGGATGTCTTCCCTCCGCGCGAGCGGTCGCGTGGT 288
Db 33955 TCGCCCGAGAGACGTGCAGATCGGCTCGATCCCGGAGCTTCGCGCGCATCGACCGG 33896
Qy 289 CCGAAGCGCGGCTCGGCGTGTTCATCCGAGGCGGCCATGTCTACGGACTGACAAACCG 348
Db 33895 GCGCGAGGCGGCTCGCCGACATCGCCGAAAGGCGAGCGGCTACGGCATCAACACGG 33836
Qy 349 GCTTCGCTCCCTTTCGAAACCGGCTGATCTCAGGTGAGATGTCCGAACCGTTCAGGCCA 408
Db 33835 GCTTCGCGCGCTTCGCGAGCACCATTCGCCACGACCACTCGAGCTCGAGTCTGAGAAGA 33776


```
QY 409 ATCTTGTCCATCATCTGGCCAGCGGCTGGGACCGGTGCTTCACTGGACGACGCGCGCG 468
Db 33775 ACCTGCTGTCTGACACGGGCTGGGCTCGGAGCCGATGGCGGCCCCGCTGTGCGCC 33716
QY 469 CCATGTTCTGCGCGCTGTGGTGTGATCGCTCAGGGAGCCTCCGCTGCGACGAGGGGA 528
Db 33715 TGTGATGGCTCAAGCTCTCAGGCTCTGAGCCTCGGCGGCGCACTCGGCACTTCTGCGCTG 33656
QY 529 CCATGCTCGCTGATGACCTGCTCAATTCGAGCTCGCTCGGCGGCTTCCAGCGCGG 588
Db 33655 TGATGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 33596
QY 589 GCACGCTGGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 648
Db 33595 GCTCGGCTGGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 33536
QY 649 GCCGGGAGACTTCTGACCGGAGCGGACCGGCTTGAACGGCGCAGAGGGCTCCGCG 708
Db 33535 GCATCGGAGACGTGTTTCATCCGCGCGA---GCGCGGAGCGCGCGGAAGGGCTGCGTG 33479
QY 709 GCGGAGGCTGCAACCGCTGATCTCTCCATCGGATGCACTGGCGGCTGCTGCTGCTGCT 768
Db 33478 TCGCGGGCTTCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 33419
QY 769 CCTCGGCATGACCGGATCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 828
Db 33418 CCGAGGCTGCAACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 33359
QY 829 GGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 888
Db 33358 CGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 33299
QY 889 CCGGCGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 948
Db 33298 AGCGCGCATCAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 33239
QY 949 GCGCGCGCTGACCGGAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1008
Db 33238 GGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 33198
QY 1009 ACGCGCGGATATCGGAGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1068
Db 33197 -----GATTGCGACAAGTGCAGGACCCGTTACAGCTGCTGCTGCTGCTGCTGCTG 33158
QY 1069 CTCGCGAGTTTCTCGGCGGCGGCTTTCGACGCTCGCATGGCATGCGGCTGCTGACGA 1128
Db 33157 AGCGCGAGTGTATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 33098
QY 1129 TCGAGCTGAAACGCTGACGCAATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1188
Db 33097 TCGAGCGAAACGCGGTATCGGACAAACCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 33041
QY 1189 TGCAGCGGGCAATTTTCATGSCCAGCATGTGGCGCTGACGCTGCGATGCGCTCGCA 1248
Db 33040 TGTGCGGGCGCAATTTTCAGCGGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 32981
QY 1249 CCGTCAACGTTCTGCGGCGCTTTCGCGAGCGCAGATTGCACTGATGACATGAAGGC 1308
Db 32980 CCGCGCGCAGATCGGCGCTTCGCGAGCGCGCATCGCTGTTGATCGAGCGGAGCG 32921
QY 1309 TGAACGCTGGGCTGCCCGCTTCTTCACCGGGGCGCGCGGCTGCTGCTGCTGCTGCT 1368
Db 32920 TCTCGGCGCTGCGCGCTTCTTCGTGAAGGAGCGGCGG-----TGAACCTCGGCTTCA 32867
QY 1369 TGGGCGCAGGTGACGCGACCGGCTCCTGGCG---AGATGCGAGCGGAGACCTG 1425
Db 32866 TGATCGGACGTGACGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 32807
QY 1426 CCTCGATCATTCGATCTTCAGAAACGCGCGCAATCAGGATGCTGCTGCTGCTGCTGCT 1485
Db 32806 CGTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 32747
QY 1486 TGGCGCGCGCTCTGCGCGGAGAGATCGACCGCTTGGGCGGAGATCCTTGGATCTCTG 1545
```

```
Db 32746 TCGCGCGCGCAAGCTTCGCGGACATCGCGGAGAACTCGGAACTCTTCGCGATCGAG 32687
QY 1546 CTCTCTGTCTTGCAAAAGCTCGGAGCTCGCTGCGGAGCGGCTAGACGGGGTGTCTC 1605
Db 32686 TGCTCGCGCGCGCAAGGCGCTCGACCTGC-----GCGCGCGCAGCAACGAGCC 32636
QY 1606 CCGCGGGAAGAGTGTGTCAGGCGCTTCGCGAGAGTTCGCGCGCTTGAGACGGAGC 1665
Db 32635 CGGCGCTCAGACACGCGATGAAGACGATTCGCGCGGAGCGTTCGCACTACGATCTCGAC 32576
QY 1666 GCGCCCTCGGACAGAAATTCGCGCTTGTACGCACT 1705
Db 32575 ACTACTTCGCGCCGACATCGCGTGTGCGCGCGCGGT 32536
```

RESULT 10
CP000031_23
WPCOMMENT

Sequence split into 41 fragments LOCUS CP000031 Accession CP000031

Fragment Name	Begin	End
CP000031_00	1	110000
CP000031_01	100001	210000
CP000031_02	200001	310000
CP000031_03	300001	410000
CP000031_04	400001	510000
CP000031_05	500001	610000
CP000031_06	600001	710000
CP000031_07	700001	810000
CP000031_08	800001	910000
CP000031_09	900001	1010000
CP000031_10	1000001	1110000
CP000031_11	1100001	1210000
CP000031_12	1200001	1310000
CP000031_13	1300001	1410000
CP000031_14	1400001	1510000
CP000031_15	1500001	1610000
CP000031_16	1600001	1710000
CP000031_17	1700001	1810000
CP000031_18	1800001	1910000
CP000031_19	1900001	2010000
CP000031_20	2000001	2110000
CP000031_21	2100001	2210000
CP000031_22	2200001	2310000
CP000031_23	2300001	2410000
CP000031_24	2400001	2510000
CP000031_25	2500001	2610000
CP000031_26	2600001	2710000
CP000031_27	2700001	2810000
CP000031_28	2800001	2910000
CP000031_29	2900001	3010000
CP000031_30	3000001	3110000
CP000031_31	3100001	3210000
CP000031_32	3200001	3310000
CP000031_33	3300001	3410000
CP000031_34	3400001	3510000
CP000031_35	3500001	3610000
CP000031_36	3600001	3710000
CP000031_37	3700001	3810000
CP000031_38	3800001	3910000
CP000031_39	3900001	4010000
CP000031_40	4000001	4109442

Continuation (24 of 41) of CP000031 from base 2300001 (CP000031 Silicibacter pomeroyi DS)

Query Match 12.3%; Score 227.2; DB 1; Length 110000;
Best Local Similarity 50.5%; Pred. No. 1.5e-20;
Matches 723; Conservative 0; Mismatches 658; Indels 51; Gaps 5;

```
QY 255 CTTGCCCTTCGCGCGCGGACCGGTCGCTCCGAGCGCGGCTCGGCGCTGTCATC 314
Db 29285 CTTGACCGGCTCGCGCGCGGTCGAGCGGCGCGGATCGCGCGCGGCG 29344
QY 315 GCGAGGCGGCCCATGTCTACGACTGACAAACCGGCTTCGTTCCCTTGGCAACCGCCTG 374
```

Db 29345 GCGGGCGATGTGCGGGTCTATGGGGTCAACACGCGGCTTTGGCAAGCTGGCCAGCTCAAG 29404
Qy 375 ATCTCAGGTGAGATGTCCGAACGCTGCAGGCAATCTGTGCCATCATCTGGCCAGCGGC 434
Db 29405 GTCCGGCGGAGGACACCGCACCTCTGACGAACCTGATCTCTGTCCCATTTGCTGGCG 29464
Qy 435 GTGGGACCGGTGTGTGATGTGACACGCGCGCGCCCATGTGTCTTGGGCGCTCTGGTGTG 494
Db 29465 GTCCGGCGCCGATCCCGCGCGCATGGCCGGCTGATGTGTGCTTTAACTGCTGTG 29524
Qy 495 ATCGCTCAGGAGCTCTCGGTGCGCAGGAGGGACCATCGCTCGCTGTGATGACCTGCTC 554
Db 29525 CTGGGGGCTGGCGCTCGGGCGTGTGGGAGCTGATCACGTTGCTTCAGGAGATGCTG 29584
Qy 555 AATTCCGAGCTCGTCCGGCGGTTCCAGCCGCGGACGGTGGCGGTGGGGTACCTG 614
Db 29585 GCGCGCATGTACCCCGGTAAATCCCGGTGAGGGCTCAGTCGGGGGCTCGGGGATCTG 29644
Qy 615 ACACCGCTTGGCATATGTGTCTGTGCTCCAGGCGCGGGAGACTTCCTGGACCGGGAC 674
Db 29645 GCGCACTGGCCATATGACGGCGTGATCATCGCGCAGCGAGCGGATGATC---AG 29701
Qy 675 GGGACGGGCTTGAACGGCGAGAAAGGCTCCGGCGCGGACGGCTGCAACCGCTCGATCTC 734
Db 29702 GGCACGGGCTCTCCCGCGCGCGAGGCGCTGGCGCGCGCTGGCTTGAACCGGATTCGCTG 29761
Qy 735 TCCCATCGCATGACCTGGCGCTGGTCAACGGGACCTCCGCCATGACGGGGATCGCGCTG 794
Db 29762 GGGCCCCAAGAGGGGCTGGCGCTTTATCAACGGCACCCAGTTCTCGACCGCATTCGCGCTG 29821
Qy 795 GTGAATGCTCACGGCTGCGCGCATCTCGGCACTCGGCGGTGGCGGTGGAGCGCCCTGCTT 854
Db 29822 GCGGCGCTGTTCGGGCGCTGGCGCGCGGACCTCTCTGCTGTGTGACGGGGCACTGTG 29881
Qy 855 GCGGAATGTCTGAGAGCCGACCGAGGCGATGGCGCGGCACTGTCCGACCTGCGGCGG 914
Db 29882 ACCGATGCGATCATGGGTGCACCGCACCCCTGCAACCGAAATCCACGCGCTGCGCGG 29941
Qy 915 CATCCCGACAGAGGAGCGCGACGAGGCTGGCGCGCGCGCTGGAGCGGACGGCGCGG 974
Db 29942 CATCGCGCGAGATCGACGCGCGCGCGGATGCGCGCTGTCTTGACGGGTCCGAGATC 30001
Qy 975 GTGTCGCGGACGCTCATTTGCGAGCGAGGCTCGACGCGCGGATATCGGACGAGCGG 1034
Db 30002 CGCAAGACACCGGAGGCGACCGG----- 30030
Qy 1035 GAGCGGGGAGGATGCTTACAGCTCGCTGCGCTCCGAGGTTCTGGGGGGGCTTC 1094
Db 30031 -----GGTGCAGGATCCCTATTGTCATCGGCTGCCAGCGGAGTGCAGGCTGCCCATG 30085
Qy 1095 GACACGCTGCATGGCATGACCGGGTGTGACGATCGAGTGAACGCGGTGACCGACAAT 1154
Db 30086 GACGTGTGGCCAGCGCGCCAGACGCTGGAGATCGAGGCAATGCGCGCACCGACAAC 30145
Qy 1155 CCGGTGTTCCGCGCGATGGCAGCGGTGCGCGCTCGACGCGGGCAATTTTCATGGGCGAG 1214
Db 30146 CCGGTGTGTGGCGGAGGGGACATATG---TCTCTGGCGCAATTTTCATGCCGAG 30202
Qy 1215 CATGTGGCGGTGACGTCCGATGCGCTCGCCACGCGCTCAACCGTTCTGGGCGGCTTGG 1274
Db 30203 CCGGTGGCTTTGCGCGCGATCTGATCGACTGCGCTGTCGAGATCGGCGCATCGCG 30262
Qy 1275 GAGCGCAGATGTGACTGTGACAGATGAAGGCTGAACGCTGGGCTGCGCCCTTCTCT 1334
Db 30263 CAACCGCGGTGGCGCTGATGGTGGATCCGACGCTGAGTTTCGATCTGCGCGCTTCTCTG 30322
Qy 1335 CACCGGGCGCGCGGCTTGAATTCGGCTTCATGGCGGACAGGTGACGGGACCGCG 1394
Db 30323 ACCCGGAGC---CGGGGCTGAATTCGGTCTGATGATCGCCGAGTCAACACCGCGCG 30379
Qy 1395 CTCCTGCGCGAGATGCGAGCGACCGGACCTGCTCGATTCGATTCGATCTCCAGAACGCC 1454

Db 30380 CTGATGAGCGAGAACAGCAATCTGGCAATCTCTGGGTCAACGACAGACCCCCACCAGC 30439
Qy 1455 GCCAATCAGGATGTGGTCTCGCTTGGGACCATCGCGCGGCGCTCTTCCCGCGAGAAATC 1514
Db 30440 GCCAATCAGAGG-----ATCAGTATCATGCGCGCGCATGGGGCGGTGCGCTGGC 30493
Qy 1515 GACCGTTGGCGGAGATCTTGGCATCTCTCGTCTCTGTGTCACAAAGCTGCGGAGCTG 1574
Db 30494 CGGATGCTCGAGAACCTGCATTACATCTCTGGCGTTCGAGCTGCTTTTTCGCGCGCAGGG 30553
Qy 1575 CGTGGCGGACGCGCTTAGACGGGGTGTCTCCCGCGGGGAAGAGCTGTGTGCGAGCCCTG 1634
Db 30554 ATCAGTTCGCGCGCGCTTTCGACGACGACCGCGCTGCAAGCTGCGGTGCGCGCTG 30613
Qy 1635 CGGAGCAGTTCCGCGCGCTTGGAGCGGACCGCGCTTGGAGAGAAATTG 1686
Db 30614 CGCGCGAGGTGCCCGCTGGAGGAGACCGTTACATGCGCGCCGAGATCG 30665

RESULT 11
AP006618_13/c
WPCOMMENT
Sequence split into 61 fragments LOCUS AP006618 Accession AP006618

Fragment Name	Begin	End
AP006618_00	1	110000
AP006618_01	100001	210000
AP006618_02	200001	310000
AP006618_03	300001	410000
AP006618_04	400001	510000
AP006618_05	500001	610000
AP006618_06	600001	710000
AP006618_07	700001	810000
AP006618_08	800001	910000
AP006618_09	900001	1010000
AP006618_10	1000001	1110000
AP006618_11	1100001	1210000
AP006618_12	1200001	1310000
AP006618_13	1300001	1410000
AP006618_14	1400001	1510000
AP006618_15	1500001	1610000
AP006618_16	1600001	1710000
AP006618_17	1700001	1810000
AP006618_18	1800001	1910000
AP006618_19	1900001	2010000
AP006618_20	2000001	2110000
AP006618_21	2100001	2210000
AP006618_22	2200001	2310000
AP006618_23	2300001	2410000
AP006618_24	2400001	2510000
AP006618_25	2500001	2610000
AP006618_26	2600001	2710000
AP006618_27	2700001	2810000
AP006618_28	2800001	2910000
AP006618_29	2900001	3010000
AP006618_30	3000001	3110000
AP006618_31	3100001	3210000
AP006618_32	3200001	3310000
AP006618_33	3300001	3410000
AP006618_34	3400001	3510000
AP006618_35	3500001	3610000
AP006618_36	3600001	3710000
AP006618_37	3700001	3810000
AP006618_38	3800001	3910000
AP006618_39	3900001	4010000
AP006618_40	4000001	4110000
AP006618_41	4100001	4210000
AP006618_42	4200001	4310000
AP006618_43	4300001	4410000
AP006618_44	4400001	4510000
AP006618_45	4500001	4610000
AP006618_46	4600001	4710000
AP006618_47	4700001	4810000
AP006618_48	4800001	4910000

AP006618_49	4900001	5010000	
AP006618_50	5000001	5110000	
AP006618_51	5100001	5210000	
AP006618_52	5200001	5310000	
AP006618_53	5300001	5410000	
AP006618_54	5400001	5510000	
AP006618_55	5500001	5610000	
AP006618_56	5600001	5710000	
AP006618_57	5700001	5810000	
AP006618_58	5800001	5910000	
AP006618_59	5900001	6010000	
AP006618_60	6000001	6021225	
Continuation (14 of 61) of AP006618 from base 1300001 (AP006618 Nocardia farcinica [F11			
Query Match 12.2%; Score 226.2; DB 1; Length 110000;			
Best Local Similarity 50.6%; Pred. No. 2e-20;			
Matches 685; Conservative 0; Mismatches 623; Indels 45; Gaps 4;			
Qy	228	GTGGCGAGCGCGCGCGGATTGTCTTCCCTCCCGCGCGCGACCGGTGCGTGC 287	
Db	65487	GTGGCGAGGAGGCGCGCGGTGCTTCCGACGAGGCGGTGGCGCGATGCCCGAC 65428	
Qy	288	TCCGAAGCGCGCGTGGCGCTGTATCCGCGAGCGCGCCATGTCTACGACTGACAA 347	
Db	65427	AGTCGAAGCGCATCGAGGCGTGGCTGCCGACCGAAGCCGCTTACGGTGTCTCGAC 65368	
Qy	348	GGCTTCGCTCCCTTGGGAAACCGCTGATCTCAGTGAGATGTCCGAAACGCTGAGGCC 407	
Db	65367	GGGTTCGCGCGGCTGGCGGTGCGCATATCCCGTCTCGAGCTCCGCAAGCAGTTCGAGCGC 65308	
Qy	408	AATCTGTCCATCATCTGCGCAGCGCGTGGACCGGTGCTTGAATGACGACGCGCGC 467	
Db	65307	AGTCTGTGCGCTGCACCGCGCGGCTTCGGGCCCGGAGTTCGAACGCGAGGTGTGCGC 65248	
Qy	468	GCCATGTTCTGGCGCTGTCTGTGTCGATCGCTCAGGAGCGCTCCGCTGCCAGCGAGGGG 527	
Db	65247	GCGCTGATGCTGCTGGGCTGTGCACCTGCGCCACCGCGCGACCGGGTGGCGCGCGAG 65188	
Qy	528	ACCATCGCTCGCTGATCGACCTGCTCAATTCGAGCTGCTCGCGCGGCTTCCAGCGCGC 587	
Db	65187	GTGGCCGCGACCTACCGCGAACTGCTCTCGCGCGGATCAACCGCGGTGGTCACGAGTAC 65128	
Qy	588	GGCAGGTGGCGGTGGCGTACCTGACACCGCTTGGCATATGCTGCTGCTGCTCCAG 647	
Db	65127	GCGAGCTCGCTGCTGGCGGACCTCGCGCGCTGGCGACGTGGCGCTGCGCGTGTCT 65068	
Qy	648	GCGCGGAGACTTCTCTGCAACCGCGACGCGGCTTTCAGCGCGCAGAAAGGCTCCGG 707	
Db	65067	GCGGAGGACCGTGGCGGATGCGCGCGGCGAACTCCGCGCGCGCGCGCGCTGGCC 65008	
Qy	708	CGCGACGGTGCACACCGCTCGATCTTCCCATCGCGATGCACTGGCGCTGGTCAACGGG 767	
Db	65007	GAGCGCGCATCGAGCGCGTTCGAGTGGCCGAGAAGGAGGCTTGGCGCTGATCAACGGC 64948	
Qy	768	ACCTCGGCATGACCGGATCGCTGTTGATGCTCAGCGCTGCCCATCTCGGCAC 827	
Db	64947	ACCACGCGCATGCTCGGCGATGCTGCTGCGCTGCCACGATTCGCGCGGTTCCTACAG 64888	
Qy	828	TGGCGGTGGGTTGACGCGCTGCTTCCGGAATGCTTCGAGAGCGCGACGAGGCGATGG 887	
Db	64887	CTGGCGGAGTGAACGCGCGGATGAGCGTGAAGCGCTGATGGGACCGCAAGGTGTC 64828	
Qy	888	GCGCGGCATGTCGACCTCGCGCGCATCCCGGACAGAGGACCGCGACGAGGCTG 947	
Db	64827	GCGCGGCATGTCGAGGCGCTGCGCGCGCACCCCGCGCAGGCGGTGGCGGCTCGAACATG 64768	
Qy	948	GCGCGCGCTGAGAGGCGCGCGGTGTGTCGCGACGCTATTTGCGACGCGAGGCTC 1007	
Db	64767	TCCCGGTGCTCGGGAATCAACCATGCTGGCCGACCGCACCGCCCGGACTGCA----- 64713	
Qy	1008	GACGCGCGCATATCGGACGAGCGGAGCGGCGGAGGATGCTTACAGCTCGCGTGC 1067	
Db	64712	-----CCGTGGTGCAGGACGCTTACTCGCTGCGCTGC 64681	

Qy	1068	GCTCGCAGGTTCTCGGGCGGGCTTCGACACGCTCCGATCGCATGCCGGTGTCTGACG 1127	
Db	64680	GCGCGCAGGTTGCGCGCGCGCGCGAGGACACCTTCGCGCACCGCGAAACGGGTTCGCGAG 64621	
Qy	1128	ATCAGAGTGAACCGGTTGACCGCAATCCGCTGTTTCGCGCCGATGGCAGCGTTCGCCGCC 1187	
Db	64620	TGGGAGCTGGCGCGCGCGGTGACACACCGCTGTGAGCGTTCGACGCGCGGTGGAATCC 64561	
Qy	1188	CTGACGCGGGCAATTTTCATGCGGCGAGATGTGGCGGTGACGTCGCGATGCGCTGCCACG 1247	
Db	64560	AAC-----GGCAACTTCCACGCGCGCGCGTGGCTACGTCGCTGGAATTCCTCGCCATC 64507	
Qy	1248	GCCCTACCGTTCTGGCGGGCTTGGCGAGCGCGAGATTGACGCTCTCACAGATGAAGG 1307	
Db	64506	GTGTGCGCGAGGTGGCGAGCATCAGCGAACCGCCACCGCATGTTCTCTGACAGCGG 64447	
Qy	1308	CTGAACCGTGGGCTGCGCCCTTCTTCCACGCGCGCGCGCGGTTGAAATTCGCGCTTC 1367	
Db	64446	GCGAGCCAGGCTTGGCGCCCTTCTTGGCGACGACCCCGCGG---TGACAGCGGGCAC 64390	
Qy	1368	ATGGGCGACAGGTGACGCGGACCGCGCTCTTGGCGGAG---ATGCGAGCCACGAGGACT 1424	
Db	64389	ATGATCGCCAGTACACCCAGCGCGCATGCTCTCCGAGCTCAAGCGCTGCGCGTCCGCC 64330	
Qy	1425	GCCTCGATCATTTGATCTCCAGAACCGCGCCCATCAGGATGTGCTCGCTTGGGACC 1484	
Db	64329	GCCAGCGTGGACTCCATCCCGTCTCTCGCGGATGACGAGGATCAGCTCTCATGGGATGG 64270	
Qy	1485	ATCCCGCGCGCTCTCCGCGAGAAAGATCAGCGCTTGGGCGGAGATCTTTCGATCTCTC 1544	
Db	64269	TCGCGCGCGCGACAGTGGCGCGGCGCTTCGACGCGCTCACCGCGCTGCGCGTGCAG 64210	
Qy	1545	GCTCTCTCTTTCGACAAAGCTGCGAGCTGCGC 1577	
Db	64209	GCGCTGACCGCGCGCGCGCTTCGAGCTGCGC 64177	

RESULT 12

STMHUTH	2946 bp	DNA	linear	BCT 26-APR-1993
LOCUS	Streptomyces griseus histidine ammonia-lyase (huth) gene, complete cds; ORF, partial cds.			
DEFINITION	M77841.1 GI:153316			
ACCESSION	M77841.1			
VERSION	histidase; histidine ammonia-lyase; histiding ammonia-lyase;			
KEYWORDS	nonoxidative deamination.			
SOURCE	Streptomyces griseus			
ORGANISM	Streptomyces griseus			
REFERENCE	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
AUTHORS	Streptomycineae; Streptomycetaceae; Streptomyces.			
TITLE	Wu, P. C., Kroening, T. A., White, P. J. and Kendrick, K. E.			
JOURNAL	Purification of histidase from Streptomyces griseus and nucleotide			
PUBMED	Sequence of the huth structural gene			
COMMENT	J. Bacteriol. 174 (5), 1647-1655 (1992)			
FEATURES	1537807			
source	Original source text: Streptomyces griseus (library: lambda EMBL4; NRRL-B-2682) DNA.			
CDS	Location/Qualifiers			
	1..2946			
	/organism="Streptomyces griseus"			
	/mol_type="genomic DNA"			
	/db_xref="taxon:1911"			
	/tissue_lib="lambda EMBL4; NRRL-B-2682"			
	<1..727			
	/note="ORF upstream of huth"			
	/codon_start=2			
	/transl_table=11			
	/protein_id="AAA26768.1"			
	/db_xref="GI:153317"			
	/translation="IVLHRAWGELYVARQGPVFRADADFATVLAAVASGIAOT			
	ERLVEVRKLAFTDPLTGLANRAVDVRLDEAMERHVDATVSVSLVCDLNLKAVNDT			

BA000040_07 700001 810000
BA000040_08 800001 910000
BA000040_09 900001 1010000
BA000040_10 1000001 1110000
BA000040_11 1100001 1210000
BA000040_12 1200001 1310000
BA000040_13 1300001 1410000
BA000040_14 1400001 1510000
BA000040_15 1500001 1610000
BA000040_16 1600001 1710000
BA000040_17 1700001 1810000
BA000040_18 1800001 1910000
BA000040_19 1900001 2010000
BA000040_20 2000001 2110000
BA000040_21 2100001 2210000
BA000040_22 2200001 2310000
BA000040_23 2300001 2410000
BA000040_24 2400001 2510000
BA000040_25 2500001 2610000
BA000040_26 2600001 2710000
BA000040_27 2700001 2810000
BA000040_28 2800001 2910000
BA000040_29 2900001 3010000
BA000040_30 3000001 3110000
BA000040_31 3100001 3210000
BA000040_32 3200001 3310000
BA000040_33 3300001 3410000
BA000040_34 3400001 3510000
BA000040_35 3500001 3610000
BA000040_36 3600001 3710000
BA000040_37 3700001 3810000
BA000040_38 3800001 3910000
BA000040_39 3900001 4010000
BA000040_40 4000001 4110000
BA000040_41 4100001 4210000
BA000040_42 4200001 4310000
BA000040_43 4300001 4410000
BA000040_44 4400001 4510000
BA000040_45 4500001 4610000
BA000040_46 4600001 4710000
BA000040_47 4700001 4810000
BA000040_48 4800001 4910000
BA000040_49 4900001 5010000
BA000040_50 5000001 5110000
BA000040_51 5100001 5210000
BA000040_52 5200001 5310000
BA000040_53 5300001 5410000
BA000040_54 5400001 5510000
BA000040_55 5500001 5610000
BA000040_56 5600001 5710000
BA000040_57 5700001 5810000
BA000040_58 5800001 5910000
BA000040_59 5900001 6010000
BA000040_60 6000001 6110000
BA000040_61 6100001 6210000
BA000040_62 6200001 6310000
BA000040_63 6300001 6410000
BA000040_64 6400001 6510000
BA000040_65 6500001 6610000
BA000040_66 6600001 6710000
BA000040_67 6700001 6810000
BA000040_68 6800001 6910000
BA000040_69 6900001 7010000
BA000040_70 7000001 7110000
BA000040_71 7100001 7210000
BA000040_72 7200001 7310000
BA000040_73 7300001 7410000
BA000040_74 7400001 7510000
BA000040_75 7500001 7610000
BA000040_76 7600001 7710000
BA000040_77 7700001 7810000
BA000040_78 7800001 7910000
BA000040_79 7900001 8010000

BA000040_80 8000001 8110000
BA000040_81 8100001 8210000
BA000040_82 8200001 8310000
BA000040_83 8300001 8410000
BA000040_84 8400001 8510000
BA000040_85 8500001 8610000
BA000040_86 8600001 8710000
BA000040_87 8700001 8810000
BA000040_88 8800001 8910000
BA000040_89 8900001 9010000
BA000040_90 9000001 9105828
Continuation (69 of 91) of BA000040 from base 6800001 (BA000040 Bradyrhizobium japonicum
Query Match 11.7%; Score 217; DB 1; Length 110000;
Best Local Similarity 50.1%; Pred. No. 3e-19;
Matches 711; Conservative 0; Mismatches 660; Indels 48; Gaps 5;
QY 162 GCCATGAGCCCCCGAAGCCGGCGTTCGAGTCGACATGCCACATCGATCGACAGGGCC 221
DB 65945 GCAATGACGAGACGAGACGCGCGGATCTGCTCAACCGGGNAAGGTGAGCTTCACGAT 65886
QY 222 CATGCCGTGGCAGCGCGCGCGGATTTGCTTGGCCCTTCGGCGCGGACCGGTGC 281
DB 65885 CTCGCGCGGTGCTTGGCGGACAGCCGTCGTCTCGATCCGTCTTGGCGCGGTGC 65826
QY 282 CGTGGTCCGAAGCGCGGTTCGGCGTGTCTATCCGAGAGCGCGCATGTCTACGGACTG 341
DB 65825 GAAGCAGCCGCGGATCTGTCGAAGCGCGCAAGCGGACACCCCGCTCTACGGCATC 65766
QY 342 ACAACCGGCTTCGGTCCCTTTCGAAACCGCTGTATCTCAGGTGAGATGTCCGACGCTG 401
DB 65765 AATACCGGCTTCGGAAGCTCGCTCGAAGCGCATCCCGCTCCAGACCGCGCTGTCT 65706
QY 402 CAGGCCAATCTTGTCCATCATCTGGCCAGCGCGGTGGACCGGTGCTTGAATGACACGACG 461
DB 65705 CAGCGCAACCTCATGCTCGCATTTGCTGCGTGTGCGGTCCGGCGACCGCGGACCGATC 65646
QY 462 GCGCGCGCATGTTCTGGCGCGTCTGGTGTGATTCGCTCAGGAGAGCTTCGGGTGCCAGC 521
DB 65645 GTGCGGCTGATGATGGCGCTGAAGATCATCTCGTGGGCGCGGCATCTGGGTACGC 65586
QY 522 GAGGGACCATCGCTCGCTGATCGACTGCTCAATTCGAGAGCTCGCTCCGCCGCTTCCC 581
DB 65585 CGCAGAGTATCGAGCAGTTTGACAGGAATGCTGGCGCGCGCTCTGTCCGCTGTCCG 65526
QY 582 AGCCGCGCACGCTGGCGGTCCGGTGACCTGACACCGCTTGGCGCATATGTTGCTCTGC 641
DB 65525 CAGCAGGGCTCGTTCGGCGCTTCGGTGACCTCGCGCGCTCGCGCATATGACCGCGCTC 65466
QY 642 CTCAGGGCCGGGAGATTCTCTGGACCGGACCGGACCGGGCTTGACGGCGCGAAGGG 701
DB 65465 ATGATCGCGAGGACAGGGGATTGTC--GACGGCAAGACCGTGTCCGGCGCGAGGCA 65409
QY 702 CTCGGCGCGGACGGCTGCAACCGCTCGATCTCTCCATCGCGATGACTGGCGCTGGTC 761
DB 65408 CTGCTCCGCCCGGGCTTGGCGCGCTGACCTTCGGGCGCAAGGAGGGGCTCGCGCTGATC 65349
QY 762 AACGGGACCTCCGCGCATGACCGGATCGCGTGGTGAATGCTCACGCTGCGCGCATCTC 821
DB 65348 AACGGCACAGTTTTCGACCGGCTATGCCATTTCCGGCGGTGCTGGTGGTTTCATCTG 65289
QY 822 GGCAACTGGCGGTGGGTTGAACCGGCTTGTTCGGAATGTCTTGAGAGCGCGGACCGAG 881
DB 65288 GCCCGTCCCGCTCGTCAACCGGCGCTGTTCGGTTCGATGGCGGATGGGCTTCCACAGC 65229
QY 882 GCATGGGCGCGGACATGTTCGACCTCGGCGGATCCCGGACAGAGGACCGCGAGCG 941
DB 65228 CCGTTCGCCCTCGAAATCCAGGGCGCTGCGGGGCCATGCGGGGAGATGCGCGCGGTGCG 65169
QY 942 AGGCTGCGCGCGCGGTGGACGCGCGGCTGCTCCGGGACGCTCATTTGCCGAGCGG 1001
DB 65168 AGCTGACCGCGCTGCTCGACCGGACGACATCCGCTGTCTGCGCATCTCGAAGGCGA---- 65113

Qy	1002	AGGCTCGACCCGCGCGATATCGGGA	CGGAGCCGGAGCGGGGAGGATGCTTACGCGTG	1061
Db	65112	-----	CGAACCGCTGCAGGATCCTATTGCGCTG	65085
Qy	1062	CGCTGGCTCGCGAGGTTCTCGGGCGGGCTT	CGACACGCTCGCATGGCATACCGGGTG	1121
Db	65084	CGTGCAGCGCGAGGTCGCGAGGGCGCGCTCGAT	CTGATCATCGACGAGATGCGCGAGCC	65025
Qy	1122	CTGACGATCGAGCTGAACCGCGGTGACCGA	CAATCCGGTGTTTCGCGCCGATGCGCAGCGTG	1181
Db	65024	CTGATCTCGAGGCCAATGCGGTACCGACA	CAACCCGCTCGTCTGTCGAGACGGCGGAG	64965
Qy	1182	CCGCGCTCGACCGGGGCAATTCATGGG	CCAGCATGTGCGCTGACGTCCGATGCGCTC	1241
Db	64964	ATCGTCTC---CGCGCGCAATTCACCG	CGAGCCGCTGGCGTTCGCGCGCATACGATC	64908
Qy	1242	GCACGCGCGTACCGTTCTGGCGGGCTT	GCGAGGCGCCAGATTGACATCTCGACAGAT	1301
Db	64907	GCGCTCGCGTGTGCGAGATCGGCGCGAT	CGGCGCGATCGAGCGCGCATCGCGACGCTGGT	64848
Qy	1302	GAAGGCTGAACCGTGGGCTGCCCTTCT	CACCGGGGCCCGCGCGGTTGAATTCC	1361
Db	64847	CCGCGCTGAAATTCGTTCTGCGCCGTT	TCTCAGGCCGATCCCGCG---TCAATTCC	64791
Qy	1362	GGCTTCATGGGCGCACAGGTGACGGG	ACCGCGCTCTCGGCCGAG---ATCGAGCCACG	1418
Db	64790	GGCTTCATGATCGCGAAGTGACGGG	CGCGCTCTATCGCGAGACAAGCAGCGGTGCG	64731
Qy	1419	GGACCTGCTCGATCCATTGATCTCACA	GAAAGCCGCCAATCAGGATGCTGCTCGCTT	1478
Db	64730	GCTGCTGCTCGATTCGATTCGAGCC	ACACGCGCCCAATCAGGAAGACCATGTGT	64671
Qy	1479	GGGACCATCGCGCGCTCTGCGCG	GAGAAGATCGACCGTTGGCGGAGATCCTT	1538
Db	64670	GCGCGCATGCGCGCGCGCTCGCCGAT	ATGCGGACAAATCTCGCGCCATCTCGCG	64611
Qy	1539	ATCTCGCTCTGTTCTTGCACAA	GTGCGAGTGTGCG 1577	
Db	64610	ATCGAGCTTCTGTCGCGCGCAAG	CGCATTACGCTGTGCG 64572	
RESULT 14				
AF319998	49736 bp	DNA	linear	BCT 28-MAY-2001
LOCUS				
DEFINITION	Stigmatella aurantiaca myxalamid biosynthetic gene cluster,			
ACCESSION	complete sequence.			
VERSION	AF319998			
KEYWORDS	AF319998.1 GI:14210834			
SOURCE	Stigmatella aurantiaca			
ORGANISM	Stigmatella aurantiaca			
REFERENCE	Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;			
AUTHORS	Cystobacterineae; Cystobacteraceae; Stigmatella.			
TITLE	1 (bases 1 to 49736)			
JOURNAL	Silakowski,B., Nordsiek,G., Kunze,B., Blocker,H. and Muller,R.			
PUBMED	Novel features in a combined polyketide synthase/non-ribosomal			
REFERENCE	peptide synthetase: the myxalamid biosynthetic gene cluster of the			
AUTHORS	myxobacterium Stigmatella aurantiaca Sgal5			
TITLE	Chem. Biol. 8 (1), 59-69 (2001)			
JOURNAL	1182319			
PUBMED	2 (bases 1 to 49736)			
REFERENCE	Silakowski,B., Nordsiek,G., Bloecker,H. and Mueller,R.			
AUTHORS	Direct Submission			
TITLE	Submitted (07-NOV-2000) MX, GBF, Mascheroderweg 1, Braunschweig			
JOURNAL	38124, Germany			
FEATURES	Location/Qualifiers			
source	1. .49736			
	/organism="Stigmatella aurantiaca"			
	/mol_type="genomic DNA"			
	/db_xref="taxon:41"			
	complement(2..1306)			
	/notes="ORF1"			
	/codon_start=1			
CDS				

	/transl_table=11	
	/product="putative aspartokinase"	
	/protein_id="AAK57182.1"	
	/db_xref="GI:14210835"	
	/translation="MRPIGSAQAGRGTSALALIVOKYGGTSVGDTERLKNVAKRCIA	
	AQAGHDVVVVSAMSGETNRLLKLVSQITERNEREQDVVATGEQVSIGLVALAIO	
	TQRRKAVSLGHQVRIITDSTFSKARIKSIDAERIVDALKKNIIVVAGFGQDGEHN	
	VTLRGSGDITVALAALAKADACEIYTDVGVYTTDPNMPVPAARAKRIERYEMLJE	
	LASGAKVQIRSVFAMKYKPLVWKFSITDDPGTGLVCEDAAMEDEVLSYATKRN	
	EAKIARIGVDPVGAAKIFGALDEQSIIVDDLIIVQNSKDGRTDVTFTVGKADLTAK	
	DVVKTKSVKAEVETDDOVKSVIGVGRNHSVAAKMTFTVLAGAGNVVQWISTIS	
	EIKVSCVHSNTELAVRSLHTAFGLDKPEAGAVSEVSLKG"	
	1363..2895	
	/notes="ORF2"	
	/codon_start=1	
	/transl_table=11	
	/product="putative histidine ammonium lyase"	
	/protein_id="AAK57183.1"	
	/db_xref="GI:14210836"	
	/translation="MSRPRLNIDGDTLKEELIQVARHTVTVELAPAAARVKAARDL	
	VDRVAGDTPSYGINTGFTGLAEVRIDKKDLRELQNLILSHAAAGVSGPLPEARVL	
	LILRCNVLAKGYSGIRPETLALALEMLNRDVPVPERGSGAGDLAPLHALLVPI	
	CGGEAPYKGERLPAAQALERAGLKPVLEAKGLALVNGTOAMCAVGTLLQRAEMLA	
	DIADLAGMTLGLGSHKPFIPETQVRAHEGQKCAAHRELLDADSALVESVNGS	
	KVQDPYSLRMPQVHGAREGLSFARRILLEINSATDNPLVFETERIVVSGNFHQ	
	PVSLADVAAMALTQLSAISERRVEQLVNPALSGLPFLAKNSGLNSGMIAQVTSAA	
	LVAESVLSHPASVDSIPSSAGREDHVSMTAALKGROVADFTRSCLATELIVAAQA	
	LDYQRPTRAGKGPAAVELIRSKIPTMEKDRELHRDIAAVALIDSGELLNAVRTATR	
	GQ"	
	complement(3275..7822)	
	/genes="mxA"	
	complement(3275..7822)	
	/genes="mxA"	
	/note="non-ribosomal peptide synthetase; contains a	
	condensing, adenylation, and PCP reductase domains"	
	/codon_start=1	
	/transl_table=11	
	/product="MxA"	
	/protein_id="AAK57184.1"	
	/db_xref="GI:14210837"	
	/translation="MSVRLLTTLAAKGVRLLEGDKLOVEASKAVLTPELRAELSAN	
	KADILFLRKHSRETSBELGQPLARPEVILPSSQERLWFLDRLSPGTSQYNVHLG	
	LSRAGDLDTKALRRSLDELVRREHLRTHFFVEGSPQVIVSPDTGVALTIIVELPGL	
	TSQRREETIYRRSDELAQQFDLAKGPLLRVTVALGPDFFALFVTKHHIITDWSLIG	
	VFLRMLYASFTSVRGAALPVSMQFADSLRERAWLAGESGARERAWKGLKGL	
	PLQLPVDHAVSTSHRGATVPLALSALSDALRELASRGCSFTVLPAATSAHLH	
	RYSGQVDFGVTVIANRGSVPADLIGFIANTLALRCDLGEPTFSOWLARHVVLEA	
	LDYQAPFSEVVQTVGASRDGGLNPLVRACFTLESIPAPTLIDLPGTSSWFLNAPDGS	
	VEGAKFELSLLAASEKGLAGMLEYSREVFDASTVERVMVGHFQVILLESIVAHEVPL	
	SKLPLLATABERGRLNWNQVPLDVPALCMHELVAQAERTPOAVVVSQKTLTYAB	
	LNRRANQLAHLRLRIGIQKEERVGLCVERTEDIVIGLLAILKAGAVYPLDPAYPEK	
	YLITYSGTGKPKGVMIEHRNAVAFILWMSVPKELAGTASTSICPDLSPATPEAIF	
	PLCCGAKVIAKNALSELPEAPAEVILINVPFAMGALLRSGCVSSVALNLAGEA	
	LAGALVDQIYQLDHDVDFNVLNPGSETTYSITFVRNNGQPTTIGRPVGNQTYALDS	
	NREPMFIVGVEYVIGMGVARGYLGRPELTAEFRVRFPGGSEARLYRTGLARWL	
	PDQLRYLGRMBDHQVLRGFRIBLGEIGAVLMEHSIGIRDAVVVRVVRGLGADKQLVAV	
	VEGEKAPEAEIRDYLKSLPEYMPVFLFVGLDALPLTPNGKVDRAALPAPERHSG	
	PAKEHAVPTPGEESLAAIWRQVIGVEQIGAHNFPELGGHSLLYRVVLARSASCA	
	DIPRLALQATLLEENARAVEAATGSLPAHDVTVEEADADLDAIEALGKALPVVVG	
	ALRTILLTGATGFLGAPLLEECRRTDARIYCLVRSKTQEGNMNRINKNLESYLMNE	
	ALAPRIYVPRGDTGQPLGLSEKFORLSEEDAIYHNGALVNFLPYEYSRMANVIG	
	TRILRLATRTRIKPLHYVTSVSLPLGRKAPIREDEPGLGESSLGGVGAQWVAEK	
	LYPEASRGI.PVTILRPGVTHSGRTGAWNTDLCVRLTKGCVRMGVAPSDALLDIT	
	PDYVSSAIVDLSMRPESICQTHLYNPQFVRADENMYMRAFGYGLRVLVYDQWLS	
	LGSAASSDSELGLMLFLQGPPEPDRSVGGPRMVVCDSDGTLKALGGTGTSCPSVDAS	
	LISTYLSLIVHRRGFLKAPEVR"	
	complement(7819..11520)	
	/genes="mxA1"	
	complement(7819..11520)	
	/genes="mxA1"	
	/note="PKS; ketosynthase; acyltransferase"	
	/codon_start=1	

Qy	868	GAGGCGGACCGAGGCA	TGGGCGCGGCGCA	CTGTCCGACCTTGGCGCCGCGATCCCGGACAGA	927
Db	2060	TGGGCGAGCACAAGCCCTTCA	TTCCGCGAGATCCAGGACGTGCGCGCGCCACA	CGAAGGCCGAGA	2119
Qy	928	AGGACGGCGGACGAGGCTGCGCGCCCGCTGGACGCGCAGCGCGCGGTGTGTCCGGCAGC	987		
Db	2120	AGGCTGGCGCGCCACTTGGCGGAGTTGTGGCGGACAGCGCACTGTTGGAGAGCCACG	2179		
Qy	988	TCATTTGGCGAGCGAGGCTCGACGCCCGCGCATATCGGGA	CGGAGCCGAGCGCGGGCAGG	1047	
Db	2180	T-----	-----GAACTGCAGCAAGGTC	CAGG	2200
Qy	1048	ATGCCTACAGCTCGGCTCGCCTCGCAGAGTTCTCGGGGGGGTTC	CGACACGCTCCGCAT	1107	
Db	2201	ACCCCTACAGCTCGGCTGCATGCCCAAGTGCACGGCGCGCGCGCGAGGGGTGTCTCT	2260		
Qy	1108	GGCATGACCGGGTGTCTACGATCAGAGCTGAACGCGGTGACCGACAATCCCGTGTTTCCGC	1167		
Db	2261	TTGCGCGCGCATCTTGGAAGTGAATCAACAGCGCTACGGACAACCCCTGTGT--GT	2317		
Qy	1168	CCGATGCGAGGTGCCGCCCTGCACGGGGGCATTTTCATGGGCCACAGATGTGGCGGTGA	1227		
Db	2318	TCGTGGAGACGGAGCGCATCGTCTCGGGCGGCAACTTCATTGGCCACGCCGTGTCTCTCG	2377		
Qy	1228	CGTCCGATCGCTCCGACGCGCCGTCAACCGTTCTGGCGGGGCTTTGCGGAGCGCCAGATG	1287		
Db	2378	CGCTGGACGTGGCGCCCATGGCGCTCAGGCNACTGTGGCCCATCAGCAGGCGCGCGTGG	2437		
Qy	1288	CACGCTTGACAGATGAAGGCTGAACCGTGGGCTGCGCCCGCTTCCCTCCACCGGGGCCCGG	1347		
Db	2438	AGCAACTCGTCAACCCAGCGCTGTGGGGGTGCGCGCGT-----TCCTTGCGAAGAAT	2491		
Qy	1348	CCGGGTTGAATTCGGGTTTCATGGGCGCACAGGTGACGGCGACCGCGCTCTCTGGCCGAGA	1407		
Db	2492	CGGGGCTCAACTCGGGGTTTCATGATGCCCAGGTGACGAGCGCGCGCTGGTGGCCGAAT	2551		
Qy	1408	TGCAG--CCACGGGACCTGCCTTCGATCCATTCGATCTCCAGAACCGCGCCCAATCAGG	1464		
Db	2552	CCCGCGTGCTCAGCCACCCCGCTCGTGGACTCCATCCCTCTCCGAGGGCGAGAGG	2611		
Qy	1465	ATGTGCTCTCGCTTGGAGACCATCGCGCGCGCTCTGCGGGAGAGAAATCGACCGTTGGG	1524		
Db	2612	ACCACGCTTCATGGGATGACACGCGCTCAAGGGCCGCTCAGGTGGCCGACTTCACCC	2671		
Qy	1525	CGGAGATCCTTGCGATCTCGCTCTCTGTCTTGACAAGCTGCGGAGCTGCGTGC	CGGCA	1584	
Db	2672	GCTGCTGCTCGCCATCGAATGCTGTGTGGCGGCCAGGCACTGGACTACGCGCAGCCCA	2731		
Qy	1585	GCGGCTTAGACGGGGTGTCTCCCGCGGGGAAGAGCTGTGGTGCAGGCGCTTGGCGGAGCAGT	1644		
Db	2732	CCCGGGGGGAAAAAGGCCGAGCGCGGTACAGAGCTCATCCGAGCA-----AGA	2782		
Qy	1645	TCCCGCGCTTGACACGACCGGCCCTTGGGACAGGAAATTCGCGCGCTTGTACTCGCAC	1704		
Db	2783	TTCCACATGGAAGAGACCGGAGCTGCACCCGGGATATCGCAGCGCTGAGCGCGCTCA	2842		
Qy	1705	TC	1706		
Db	2843	TC	2844		

[illegible]

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 10, 2005, 11:46:35 ; Search time 7538 Seconds
(without alignments)
11482.635 Million cell updates/sec

Title: US-10-621-826-2
Perfect score: 1850
Sequence: 1 ccgagcgccatccacgag.....tcagctgatcccgacgagg 1850

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_hic.*
5: gb_est4.*
6: gb_est5.*
7: gb_est6.*
8: gb_est7.*
9: gb_gss1.*
10: gb_gss2.*
11: gb_gss3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	309.8	16.7	400	9 B07840	B07840 8536T101C31
2	122.2	6.6	2151	10 CL972100	CL972100 OsIFCC041
3	119	6.4	2598	4 AY103647	AY103647 Zea mays
4	111	6.0	4590	10 CL982033	CL982033 OsIFSC046
5	104.8	5.7	1740	10 CL967485	CL967485 OsIFCC040
6	102.2	5.5	669	9 BZ396558	BZ396558 EINAG81TF
7	97	5.2	812	10 CG301209	CG301209 OG3AG33TV
8	97	5.2	845	10 CG362861	CG362861 OG3A210TH
9	97	5.2	888	9 CG664163	CG664163 OGNAF84TV
10	94.4	5.1	737	7 CF842933	CF842933 pSHB023XA
c 11	92	5.0	788	8 DN640249	DN640249 UMC-bend
c 12	92	5.0	930	8 CX949440	CX949440 UMC-bcl_0
13	90.6	4.9	847	8 DR785713	DR785713 ZM BFB000
14	90.4	4.9	786	6 CB653003	CB653003 OSJNEC03J
15	90.4	4.9	866	6 CB642693	CB642693 OSJNEB03A
16	90.2	4.9	856	6 CB654478	CB654478 OSJNEC07A
17	90	4.9	841	10 CG298372	CG298372 OGDNR64TV
18	90	4.9	1429	9 BZ578760	BZ578760 msh2_5990
19	89.8	4.9	831	9 CG675672	CG675672 OGUES16TV
20	89.6	4.8	808	6 CB655968	CB655968 OSJNEC09M
21	89.2	4.8	859	6 CB632830	CB632830 OSJIEB11P
22	88.4	4.8	796	6 CB633009	CB633009 OSJIEB06N

c 23	88.2	4.8	662	10 CL191959	CL191959 104 414_1
24	88	4.8	667	6 CA122223	CA122223 SCJFIR101
25	88	4.8	766	7 CN139042	CN139042 OX1 15_D0
26	88	4.8	794	6 CB649950	CB649950 OSJNEB14C
27	87.8	4.7	614	6 CA178525	CA178525 SCRFST104
28	86.8	4.7	778	6 CB632619	CB632619 OSJIEB10P
29	86.4	4.7	857	6 CB627358	CB627358 OSJIEB02E
c 30	86.4	4.7	917	10 CZ399499	CZ399499 ZMMBFO180
31	85.6	4.6	611	7 CO969005	CO969005 BeE9ON26F
32	85.4	4.6	773	6 CB670025	CB670025 OSJNEE020
33	85.2	4.6	597	6 CA132692	CA132692 SCEQRT102
34	85.2	4.6	805	6 CB651372	CB651372 OSJNEB16E
c 35	85.2	4.6	824	10 CZ539004	CZ539004 SRAA-sad2
36	84.8	4.6	764	6 CB632150	CB632150 OSJIEB10D
c 37	84.2	4.6	680	10 CM163154	CM163154 104 570_1
38	84.2	4.6	691	10 CW307776	CW307776 104 796_1
39	84.2	4.6	812	6 CB642197	CB642197 OSJNEB02E
40	84	4.5	817	6 CB666404	CB666404 OSJNEB13E
41	83.6	4.5	427	7 CO962233	CO962233 BEEL20N23
42	83.4	4.5	909	6 CB629973	CB629973 OSJIEB06K
43	83.2	4.5	658	6 CA178264	CA178264 SCRFST104
44	83.2	4.5	803	6 CB633242	CB633242 OSJIEB11P
45	83.2	4.5	2142	10 CL967487	CL967487 OsIFCC040

ALIGNMENTS

RESULT 1
B07840/c
LOCUS
DEFINITION
B07840 400 bp DNA linear GSS 15-NOV-1997
8536T101C3120194T3 Rhodobacter sphaeroides 2.4.1 genomic DNA
library Rhodobacter sphaeroides genomic clone 8536T101C3120194T3
similar to P42357 HUTH HUMAN HISTIDINE AMMONIA-LYASE (HISTIDASE)
Homo sapiens, genomic survey sequence.
ACCESSION
B07840
VERSION
B07840.1 GI:2055596
KEYWORDS
GSS.
SOURCE
Rhodobacter sphaeroides
Rhodobacter sphaeroides
Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
Rhodobacteraceae; Rhodobacter.
REFERENCE
1 (bases 1 to 400)
AUTHORS
Choudhary, M., Mackenzie, C., Nereng, K., Sodergren, E., Weinstock, G.M. and Kaplan, S.
TITLE
Low-resolution sequencing of Rhodobacter sphaeroides 2.4.1T:
chromosome II is a true chromosome
JOURNAL
Microbiology 143, 3085-3099 (1997)
PUBMED
9353914
COMMENT
Contact: Mackenzie, C.
Department of Microbiology and Molecular Genetics
The University of Texas Health Science Center
6431 Fannin St., Houston, TX 77006, USA
Tel: (713) 794-1742
Fax: (713) 794-1782
Email: mackenzet@utmsi.med.utmc.edu
Seq primer: pBluescript SK (-) T3
Class: shotgun.
FEATURES
Location/Qualifiers
source
1..400
/organism="Rhodobacter sphaeroides"
/mol_type="genomic DNA"
/strain="2.4.1T"
/db_xref="taxon:1063"
/clone="8536T101C3120194T3"
/lab_host="E. coli S17-1"
/clone_lib="Rhodobacter sphaeroides 2.4.1 genomic DNA library"
/notes="Vector: pLA2917; Genomic DNA from Rhodobacter sphaeroides was prepared and partially digested with Sau3AI. Size selected (20kb) fragments were subcloned into the BglII site of the cosmid vector pLA2917 (Allen, L. N. and R. S. Hanson. 1985. Journal of Bacteriology,

161:955-962. The library was then ordered around chromosome II of Rhodospirillum rubrum (Choudhary et al., 1994. Journal of Bacteriology, 176:7694-7702). The cosmids were then digested with a variety of restriction enzymes (BamHI, EcoRI, BglII, PstI, EcoRV, NotI and DNaseI) and restriction fragments subcloned into the respective multiple cloning site sites of pBluescript SK (-). Note BglII fragments were subcloned into the pBluescript BamHI site. DNaseI fragments were subcloned into the EcoRV site. All subclones were transformed into E. coli XL1Blue MRF'. All fragments were then sequenced and the sequences where possible were assembled using the GCG program GELASSEMBLE."

ORIGIN
Query Match 16.7%; Score 309.8; DB 9; Length 400;
Best Local Similarity 89.0%; Pred. No. 3.3e-56;
Matches 333; Conservative 0; Mismatches 38; Indels 3; Gaps 1;
QY 545 CGACCTGCTCAATTCGAGCTCGTCCGGCCGTTCC---AGCCGGCGCACGGTGGGCGC 601
DB 375 CGACCTCTCAANNCCGAGATCCCTCCGGCCCGTTNNAANNAACGGAAGGTGGGCGC 316
QY 602 GTCGGGTGACCTGACACCGCTTGGGATATGTTGCTCTGCTCCTCAGGGCCGGGAGACTT 661
DB 315 GTCGGGTGACCTGACACCGCTTGGGATATGTTGCTCTGCTCCTCAGGNCCTGGGAGACTT 256
QY 662 CTTGGACCGGCGGAGCGGGCTTGACGGCGCAGAGGGCTCCGGCGGCGGCTGCA 721
DB 255 CTTGGACCGGCGGAGCGGGCTTGACGGCGCAGAGGGCTCCGGCGGCGGAGCTCTCCA 196
QY 722 ACCGCTCGATCTCTCCATCGATGACATGCGGCTGCTCAAGGGACCTCCGCCATGAC 781
DB 195 ACCGCTCGATCTCTCCATCGATGACATGCGGCTGCTCAAGGGACCTCCGCCATGAC 136
QY 782 CGGATCGCTGCTGATGCTCAGCTTCCGCGCATCTCGGCACTGCGGCGGTGGGCTT 841
DB 135 CGGATCGCTGCTGATGCTCAGCTTCCGCGCATCTCGGCACTGCGGCGGTGGGCTT 76
QY 842 GACGGCTCTCTGCGGAATGCTGAGAGCGGACCGGAGGATGGCGGCGGCTGTC 901
DB 75 GAGNCCCTCTGCGGAATGCTGAGAGCGGACCGGAGGATGGNCCGCGGCACTGTC 16
QY 902 CGACCTGCGGCGCGC 915
DB 15 CGACCTGCGNCCG 2

RESULT 2
CL972100 2151 bp DNA linear GSS 21-SEP-2004
LOCUS OsIFCC041692 Oryza sativa Expressed Sequence Tag (EST) (indica
DEFINITION cultivar-group) genomic, genomic survey sequence.
ACCESSION CL972100
VERSION 1
KEYWORDS GSS.
SOURCE CL972100.1 GI:52398729
ORGANISM Oryza sativa (indica cultivar-group)
Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 2151)
Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M.,
Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L.,
Wong, G.K.S., Deng, X.W. and Wang, J.
An analysis of transcriptional regulation of the rice genome and
its comparison to Arabidopsis
Unpublished (2004)
Contact: Chen Chen
Department of Bioinformatic
Beijing Institute of Genomics
Chinese Academy of Sciences, Beijing 101300, China
Tel: 86-10-80481559

Fax: 86-10-80488676
Email: chenchen@genomics.org.cn
Rice genomic sequence.
Class: exon-trapped.
FEATURES
source 1.2151
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="genomic DNA"
/db_xref="taxon:39946"
/clone_lib="Oryza sativa Expressed Library"
/notes="Oryza sativa exon trapped genomic sequences"
ORIGIN

Query Match 6.6%; Score 122.2; DB 10; Length 2151;
Best Local Similarity 47.5%; Pred. No. 1e-15;
Matches 660; Conservative 0; Mismatches 678; Indels 51; Gaps 8;
QY 219 GCCCATGCGTGGCGAGCGCGCGGATGTTGCTTGGCCCTCGCGCGCGACCGG 278
DB 211 GCCGGCGCGCGAGCGCGCGGATGTTGCTTGGCCCTCGCGCGCGAGCGC 270
QY 279 TGCCTGCGTCCGAAGCGCGGCTCGGCGTGTCTATCGCGAGCGCGCCATGTACGGA 338
DB 271 GTCAAGGCCAGCAGCTGGGTGATGAACAGCATGGCGAACCGCACAGCTAGCGC 330
QY 339 CTGACAAACCGGCTTCCGTTTGGGAAACCGCTGATCTCAGGTGAGATGTCCGAACG 398
DB 331 GTCAACACCGGTTCCGCGCCAGCTGCACCGCGGACCAAGAGGCGCGCTGCAG 390
QY 399 GTGACGGCAATCTTGTTCATCATCTGG-----CCAGCGCGTGGGACCGGTG 446
DB 391 CGGAGCTCATAGGTTCTCAACGCGCGGCTTCCGCGACCGCGCGCGACCGCAGTC 450
QY 447 CTTGATGACGAGCGCGCGGCTTCTGGCGGCTGTGTGTCGATCGCTCAGGGA 506
DB 451 CTCGCCGCGGTGCAACGCGCGGCGGCTGCTCGTCCGGATCAACACCTGCTGCAAGGC 510
QY 507 GCCTCGGTGCCAGCGAGGCGGCTCGCTGCTGATCGACCTGCTCAATTCGAGCTC 566
DB 511 TACTCCGGCATCGGTTTCAGATCTCGAGGCGGTGCGCAAGCTGCTCAACGCAAGCTC 570
QY 567 GCTCCGGCGTTCCAGCGCGGCGACGCGTGGCGCGCTCGGCGTGAACCTGACACCGCTTGG 626
DB 571 AGCCCGTGTGCGCTCCGGGCGACGATCACCGCGTGGGCGGACCTCGTCCGCTCTCC 630
QY 627 CATATGCTGCTGCTCCCTCCAGGCGCGGAGACTTC-----CTGACCGGAGCGGAGC 680
DB 631 TACATCGCGGCTGCTCACCGCGCGGAGAACTCCGTGGGTGGCGCCGACGCGAGG 690
QY 681 CGGCTTGGCGGCGAGAGGCTTCCGCGCGGAGCGCTGCAACCGC-----TCGATCTC 734
DB 691 AAGGTGAACCGCGCGGAGGCTTCAAGATCGCGGCGATCCAGGCGGCTTCTTCGAGCTG 750
QY 735 TCCCATCGGATGCACTGGCGCTGTTCAAGCGGACCTCGCCATGACCGGATCGCGCTG 794
DB 751 CAGCCCAAGAGGCTTCGCTGTTGAAACGCGGCGGCTCGGCGCTCGGCTCTCCCTCC 810
QY 795 GTGAATGCTCAGCCTCGCGCATCTCGCAACTCGGCGGCTGGCGTTGACGGCCCTGCTT 854
DB 811 ACGTGTCTGTTGAGGCCAAGCTTCTCGCCATCTCGCGAGGTGCTCTCCGCGGTGTT 870
QY 855 GCGGAATGCTGAGAGCGCGGACCGAGGATGGCGCGCGGCACTGTCCGACCTTCGCGCGC 914
DB 871 TGGAGGTGATGAACGCGAAGCGGAGTACACCGACCACTCACCCACAAGCTGAAGCAC 930
QY 915 CATCCGCGAGAGGACCGCGGAGGCTGCGCGCGCGCTGAGACGCGGCGGCGG 974
DB 931 CACCCGGGCGAGATCGAGGCGCGGCGCATCATGGAGCACATCTTGAAGGACGCTC---- 986
QY 975 GTGGTCCGCGAGCTCATTCGCGAGCGGAGGCTCGACCGCGGCGGATATCGGAGCGGAGCG 1034
DB 987 -----GTACATGAAGGAGCGGAGAGGCT-----CGGCGAGCTCGACCCGCTGATG 1032

Db 941 TGCAGGTCATGAACCGGCAAGCCGAGTACACGACCACTGACCCACAAGCTGAAGCAC 1000
Qy 915 CATCCCGGACAGAGGACCGCGACGAGGCTGCGCGCCGCGTGGACGACGCGCGCG 974
Db 1001 CACCCGGGTCCATCGAGGCGCGCGCCATCATGAGACATCTCTGGATGGCAGCTC--- 1056
Qy 975 GTGTCCCGGACAGTCATTGGCCGAGCGGAGGCTCGACGCCGCGCATATCGGGACGAGCGC 1034
Db 1057 -----CTTCATGAACGAGCGCAAGAGTTGAACGAGCTGGACCCGCTGCTG 1102
Qy 1035 GAGCGGGGAGAGTGCCTACAGCTGCGCTGCGCTCGGAGGTTCTCGGGCGGGCTTC 1094
Db 1103 AAGCCCAAGCAGGACAGTACGCTCCGACAGCTCGCGCGAGTGGCTGGGCGCCACATC 1162
Qy 1095 GACACGCTCGCATGGCATGCCGGTCTGACGATCGAGTGAACGCGGTGACCGACAAT 1154
Db 1163 GAGTCAATCCGCGCGCCACCAAGTCCATCGAGCGAGGTCAACTCCGTGAAGCAAC 1222
Qy 1155 CCGGTGTTCCGCCGATGGCAGCTGCCCGCCCTGCACGGGGCAATTTTCATGGGCGAG 1214
Db 1223 CCGGTATGACGCTCCACCGGGCAAG--GCGCTGCACGGCGCACTTCCAGGGCACC 1279
Qy 1215 CATGTGGCGTACGCTCCGATGCGCTCGCCACGGCGCTCACCGTTCGTGGCGGGCTTGGC 1274
Db 1280 CCCATCGCGGTGTCATGGAACAACGCCCGCTCGCCATCGCCAACATCGGCAAGCTCATG 1339
Qy 1275 GAGCGCAGATTGACACCTCTGACAGATGAAGGCTGAACGCTGGGCTGCCCTTCT- 1333
Db 1340 TTGCGCGAGTTCTCCGAGCTCGTCAACAGGTTCTTACAAACAACGGGCTCACTTCAACCTG 1399
Qy 1334 --CCACCGGGCGCCCGCGGTTTGAATTCGCGCTTCATGGCGGCACAGGTGAACGGCGACC 1391
Db 1400 GCCGGACGCGCAACCCAGCCTGGACTACGGCTTCAAGGGCACCGAGATGCGCATGGCC 1459
Qy 1392 GCGCTCTGCGCGAGATGCGAGCCACCGG--ACCTGCTCGATTCATTCGATTCACG 1448
Db 1460 TCCTACTGCTCCGAGTCCAGTACTGGGCAACCCCATCAACCAACGCTGCAGAGCGCG 1519
Qy 1449 AACGCCCAATCAGGATGGTCTGCTGCTGGGACCATCGCGCGCTCTCGCGGAG 1508
Db 1520 GAGCAGCACAAACAGGACGTAACCTCCCTGGCGCTGCTCGGCGAGAAACGCGCGAG 1579
Qy 1509 AAGATCACCGTTGGCGGAGATCCTTGGCATCTCGCTCTCTGTCTTGACACAAGCTGCG 1568
Db 1580 GCGATCGACATCTGAAGCTCATGCTCTGCACCTACATCGTGGCGCTGTGCCAGGCGCTG 1639
Qy 1569 GAGCTGCGC 1577
Db 1640 GACCTGCGC 1648

RESULT 4
CL982033
LOCUS OaIFSC046542 Oryza sativa Exprim Library Oryza sativa (indica
DEFINITION cultivar-group) genomic, genomic survey sequence.
ACCESSION CL982033
VERSION CL982033.1 GI:52418552
KEYWORDS GSS.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 4590)
AUTHORS Ma,L., Wang,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
Wong,G.K.S., Deng,X.W. and Wang,J.
TITLE An analysis of transcriptional regulation of the rice genome and
JOURNAL its comparison to Arabidopsis
COMMENT Unpublished (2004)
Contact: Chen Chen

Department of Bioinformatic
Beijing Institute of Genomics
Chinese Academy of Sciences, Beijing 101300, China
Tel: 86-10-80481559
Fax: 86-10-80488676
Email: chenchen@genomics.org.cn
Rice genomic sequence.
Class: exon-trapped.
FEATURES
source 1..4590
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="genomic DNA"
/db_xref="taxon:39946"
/clone_lib="Oryza sativa Exprim Library"
/note="Oryza sativa exon trapped genomic sequences"
ORIGIN
Query Match 6.0%; Score 111; DB 10; Length 4590;
Best Local Similarity 48.9%; Pred. No. 2.8e-13;
Matches 328; Conservative 0; Mismatches 340; Indels 3; Gaps 1;
Qy 267 GCGGCGACCGGTGCGTCCGAGCGGCTCGCGCTGTCATCCGAGCGCGC 326
Db 2635 GCCTACGGCCCTATCCAGCAGAGCGTGGCTCGGTGAGCAGATCATCGCCGAAGACCGC 2694
Qy 327 CATGTCTACGGACTGACAAACCGGCTTCGCTCCCTTGCAGAACCGCTGATCTCAGGTGAG 386
Db 2695 ACGCCCTACGGCATCAACCGGTTTCGGCTCTGCTTGCACCCGCATCGCCCGTAC 2754
Qy 387 AATGTCCGAACGTCGAGCGCAATCTTTGTCATCATCTGGCCAGCGCGTGGAGACCGGTG 446
Db 2755 GATTGGAAACCTGACGCGTTCAATCTGCTGTCAACACCGCCGCGGTGGCGAGCCCC 2814
Qy 447 CTTGACTGGACGAGCGCGCGCATGTTCTGGCGGTCTGCTGTCGATCGCTCAGGA 506
Db 2815 ACCGACGACAATCTGGTGGCCCTGATGATGGTGTGAAATCAACACGCTGTTCGCGCGC 2874
Qy 507 GCCTCGCGTCCAGCAGAGGAGCCTCGCTCGCTGATCGACCTGCTCAATTCGAGCTC 566
Db 2875 TTCTCCGGCATTCGTCGGAAGTCATCCAGCGCTGATCGCGTGGTCAACGCCGAAGTT 2934
Qy 567 GCTCGCGCGCTTCCAGCGCGCGAGCGTGGCGCGCTGGGTGACCTGACACCGCTTGGC 626
Db 2935 TATCCGCATATCCGCTGAAGGGCTCGGTGCGCGCTTCCGCGACCTTGGCGCGCTGGCG 2994
Qy 627 CATATGTGCTCTGCTCCAGGCGCGGAGACTTCTTGACCGCGGAGCGGAGCTT 686
Db 2995 CACATGAGCTGTGCTGCTGGGCGAAGCCAGCGCGTTCATCAGGCGCAATGCTGCGC 3054
Qy 687 GAGCGCGCAGAGGGGTTCGCGCGCGAGCGCTGCAACCGCTCGATCTCTCCCATCGCGAT 746
Db 3055 GCGACCGAGCGCTGGCCAGGCGCG--TCTGAACCGCTGACGCTGGCGGCGAAGAG 3111
Qy 747 GCACTGCGCTGTCAACCGGACCTCGCCATGACCGGGATCGCGCTGGTGAATGCTTAC 806
Db 3112 GGCCTGGCGCTGCTGAACCGCACCCAGTCTCCGCGCGCTTCCGCTGCGCGCTGCTTC 3171
Qy 807 GCTTGGCGCATCTCGGCAACTGGGCGGTGGCTTGACGGCCCTGCTTGGCGGATGCTG 866
Db 3172 GACCGGGAAGACCTGTTGCGCCGCGGACCGTGGCGGGCAGCCTGACGCTGGGAAGCGCGC 3231
Qy 867 AGAGGCGCGACCGAGGATGGGCGCGGCGCACTGTCCGACCTGGCGCGCGCATCCCGGACAG 926
Db 3232 CTGGGCTCCCGAGCGCGTTCGATGCGCGTATCCATGAGGTGCGCGCGCGCGCGAG 3291
Qy 927 AAGAGCGCGC 937
Db 3292 ATCGAGCGCGC 3302
RESULT 5
CL967485
LOCUS CL967485 1740 bp DNA linear GSS 21-SEP-2004

DEFINITION OsIFCC040011 Oryza sativa Express Library Oryza sativa (indica cultivar-group) genomic, genomic survey sequence.
ACCESSION CL967485.1 GI:52389612
VERSION CL967485.1
KEYWORDS GSS.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Eriartoideae; Oryzaceae; Oryza.
REFERENCE 1 (bases 1 to 1740)
AUTHORS Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M., Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L., Wong, G. K. S., Deng, X. W. and Wang, J.
TITLE An analysis of transcriptional regulation of the rice genome and its comparison to Arabidopsis
JOURNAL Unpublished (2004)
COMMENT Contact: Chen Chen
Department of Bioinformatic
Beijing Institute of Genomics
Chinese Academy of Sciences, Beijing 101300, China
Tel: 86-10-80481559
Fax: 86-10-80488676
Email: chenchen@genomics.org.cn
Rice genomic sequence.
Class: exon-trapped.
FEATURES
source Location/Qualifiers
1..1740
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="genomic DNA"
/db_xref="taxon:39946"
/clone_lib="Oryza sativa Express Library"
/note="Oryza sativa exon trapped genomic sequences "
ORIGIN
Query Match 5.7%; Score 104.8; DB 10; Length 1740;
Best Local Similarity 47.7%; Pred. No. 5.7e-12;
Matches 558; Conservative 0; Mismatches 572; Indels 40; Gaps 7;
427 CCAGCGCGTGGACCGGTCTTGATGGACGACGCGCGGCCCATGGTTCTGGCGCGTC 486
19 CCGGACCGACGCGGCACGTTCTCGCGCGGAGCGCGCGCGATGCTCGTCGCGCA 78
487 TGGTGTGATGCTCA-GGGAGCTTCGGTGCCAGCAGGAGGACCATGCTGCGCTGATC 545
Db 79 TCAACACCCCTCTCCAGGGGTACTTCCGCGCATCCGTTTCGAGATCCTCGAGGCCATCACC 138
546 GACCTGCTCAATTCGAGCTCGCTCCGCGCGTTCCAGCGCGGACCGTGGCGCGTCG 605
Db 139 AAGCTGCTCAATGCGCAACGTCACGCGCGTGCCTCGCGCGGACCATCACCGCGCTCC 198
606 GTGACCTGACACCGCTTGGCGATATGCTGCTCTGCTCCAGGCGCGGAGACTTTCCTG 665
Db 199 GGTGACTTGGTCCCACTGTCTTACATTCGCGGCTCATCACCGCGCGCGAGACTCCGTTG 258
666 G-----ACCGGACGCGGACGCGGTGTGACGCGCGCAGAGGGCTCCGCGCGGACGCGTCG 720
Db 259 GCGGTGCGCGCGATGCGCGCAAGGTGACCGCGCGCGGAGGCAATCAAGATTGCGGCAAT 318
721 AACCGC-----TCGATCTCTCCCATCGCATGACCTGGCGCTGTTCAACGGACCTCC 773
Db 319 GAGCACGCGCTTCTTCGAGTTGCGAGCTTAAGGAAGGTCTTGCCATGGTGAACGCGCACGCC 378
774 GCCATGACCGGATCGCGCTGTGTAATGCTACGCGCTCGCGCATCTCGGAACCTGGCGCG 833
Db 379 GTCGGCTCTGGCTTGATGACCGGTCTCTTTGAGGCGCAAGCTCTTGCCATCTCGCC 438
834 GTGGCGTTGACGCGCTCTTTCGGAAATGTCTGAGAGCGCGGACCGAGGCAATGGCGCGCG 893
Db 439 GAGGTCTGTTCGCGCGTGTCTGTGAGGTCTATGACCGCGCAAGCGCGAGTACACCGACAC 498
894 GCATGTCTCGACCTCGCGCGCATCCCGACAGAGGACCGCGGACGAGGCTGGCGCC 953

499 CTGACACACAAGCTGAAGCACCCCTGGACAGATCGAGGCTGCGCCATCATGAGCAC 558
QY 954 CGCGTGGACGCGAGCGCGCGGTGGTCCGCGACGTCATTTCGCGAGCGGAGGCTCAGCGC 1013
Db 559 ATCTTGGAGGCGAGCTCGTACATG-----AAGCTGCGGAAGAGCTC 600
QY 1014 GCGCATATCGGAGCGAGCGCGGAGCGCGGAGGATGCTACAGCTGCGCTGCGCTCG 1073
Db 601 GCGGAGCTCGACCGCTTGTATGAAGCGGAGGAGGACAGGTACGCGCTCGGACGTCGCGC 660
QY 1074 CAGGTTCTCGGCGCGGCTTCGACACAGCTCGCATGCGCATGCGGATGCTGACGATCGAG 1133
Db 661 CAGTGGCTCGGCGCCACAGATCGAGGTATCGCTTCGCGCACCAAGTCGATCGAGCGCGAG 720
QY 1134 CTGAACGCGGTGACGCAATTCGCGTGTTCGCGCGGAGTGGAGCGTGGCCGCTTCGAC 1193
Db 721 ATCAACTCCGTCAACGACCAACCCGCTGATGATGCTCCCGTGGCAAG---CGGCTTCAC 777
QY 1194 GGGGCAATTTTCATGGGCGGACGATGTCGCTGAGCTCCGATCGCTCGCCACGCGCGTC 1253
Db 778 GGTGGCAATTTTCAGGGCGACGCCCATCGCGGTGTCATGGACCAACCGCGCTCGCCCTC 837
QY 1254 ACCGTTCTGGCGGCGCTTCGCGAGCGCGCAGATTGACGCTCTGACAGATGAAGGCTGAAC 1313
Db 838 GCTGCATCGCAAGCTCATGTTTCGCGCAGTTCTCTGAGCTCGTGAACGACTTCTACAC 897
QY 1314 CGTGGGCTGCGCGCTTCTCTCCACCGCGGCGCG---CGGGTTGAATTCGCGCTTCATG 1370
Db 898 AACGGGCTTCTTCCAACTGTCCGCTGAGGACCAACCCAGCTTGGACTATGGTTCAAG 957
QY 1371 GCGCAGAGGTGAGCGGCGACCGCGCTCTGCGCGAGATGCGAGCCACCGG---ACTGCC 1427
Db 958 GCGCGGAGATGCGCATGGCGTCTGCTCCGAGCTCCAGTTCTTGGGCAACCGCGTG 1017
QY 1428 TCGATCCATTCGATCTCCACGAACCGCGCCCAATCAGGATGCTGCTCGCTTGGGACCATC 1487
Db 1018 ACCAACCATGTCAGAGCGCGGAGCAGCAGCAACACGAGAGTGAATTCGCTTGTCTCATC 1077
QY 1488 GCGCGCGCTCTGTCGCGGAGAGATCGACCGTGTGGCGGAGATCCTTGGCATCTCGCT 1547
Db 1078 TCCTCCAGGAAGACCGCGAGGCCATCGACATCTCTGAAGCTCATGCTCTCCACGTTCTTG 1137
QY 1548 CTCTGTCTTGACAACTGCGGAGCTGCGC 1577
Db 1138 ATCGCCTGTGCCAGGCCATCGACTGCGC 1167
RESULT 6
BZ396558
LOCUS EINAG81TF EI_10.12_KB Entamoeba invadens genomic clone EINAG81,
DEFINITION genomic survey sequence.
ACCESSION BZ396558
VERSION BZ396558.1 GI:30243105
KEYWORDS GSS.
ORGANISM Entamoeba invadens
SOURCE Entamoeba invadens
REFERENCE 1 (bases 1 to 669)
AUTHORS Wang, Z., Samuelson, J., Clark, C. G., Eichinger, D., Paul, J., van Dellen, K., Hall, N., Anderson, I. and Loftus, B.
TITLE Gene discovery in the Entamoeba invadens genome
JOURNAL Mol. Biochem. Parasitol. 129 (1), 23-31 (2003)
PUBMED 12798503
COMMENT Other GSSs: EINAG81TR
Contact: Brendan Loftus
Department of Eukaryotic Genomics
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-3543
Fax: 301-838-0208
Email: enta@tigr.org
DNA was provided by Daniel Eichinger

Seq primer: TF
Class: sheared ends.
Location/Qualifiers
source
1..669
/organism="Entamoeba invadens"
/mol_type="genomic DNA"
/strain="IP-1"
/db_xref="taxon:33085"
/clone="EINAG81"
/note="Vector: pRGS2; Site 1: BstXI; Total genomic DNA was isolated from early log phase trophozoites of E. invadens IP-1 using a Qiagen plant DNA extraction kit. A shotgun medium-size plasmid library (average insert size of 10 - 12 kb) was generated by random mechanical shearing of E. invadens genomic DNA, repairing the ends of DNA fragments with T4 Polymerase, adding BstXI adaptors and ligating into the BstXI site of a pUC-derived vector pRGS2."

ORIGIN

Query Match 5.5%; Score 102.2; DB 9; Length 669;
Best Local Similarity 52.2%; Pred. No. 1.9e-11;
Matches 227; Conservative 0; Mismatches 208; Indels 0; Gaps 0;
QY 199 GCCACATCGATCTGGACAGGCCCATCCGTGGCGAGCGCGCGCGGATGTCTTGG 258
DB 230 GCAAGGTGACCTTGGACGCTGGTCCGATCCAGGCGCGGCGTTCGGCTTGGCTCG 289
QY 259 CCCCTCCGGCGCGGACCGGTGCGTCCGTCGTCGAGCGCGGCTCGGCGCTGTCTCCG 318
DB 290 ACCCGTCGGTCCGCGACGGCATGTGCGTCCGAGGCGCGCGTCCGCGCACATCGTCG 349
QY 319 AGGCGGCCCATGTCTACGGAGTACAAACCGCTTCGGTCCCTTGGCAACCGCTGATCT 378
DB 350 ACGACCAAGTGTCTACGGCATCAACACCGCTTCGGCAAGCTCCGCGACGCGGCAATCG 409
QY 379 CAGGTGAGAAATGTCGGAACGCTGCAGGCGCAATCTTGTCCATCATCTGGCCAGCGCGTGG 438
DB 410 GCAACGACCACTGGCGGAGTGCAGCGCAACCTGGTGTCTGCGACAGCGTGGGACAG 469
QY 439 GACCGGTGTTGACTGACGACGCGCGCGCGCATGGTTCTTGGCGGCTGTCTGGTTCGATCG 498
DB 470 GCGAGCGCTGGTGGCGCGCTGTGCGCATGATCTCTCGGACCAAGCGCGTGGAGCGCTGG 529
QY 499 CTCAGGAGAGCTCCGGTCCAGCGAGGAGGACCATCTGCTCCCTGATCGACTGTCTCAATT 558
DB 530 CGCGCGCCATTCGGGCTGGCGCCCGCTGTGTCGATGCGTCTGCTGCGCTGTTCACG 589
QY 559 CCGAGCTCGTCCGCGCGTTCCAGCGCGCGCACGCGTGGCGCGTCCGGTGAACCTGACAC 618
DB 590 CGGCGGTACGCGGAGCATTCGGTGCAGAGGCTCGGTGGTGGTCCGCGGACCTCGCGC 649
QY 619 CGCTTGGCATATGG 633
DB 650 CGTCCGCGCATGG 664

RESULT 7

CG301209 812 bp DNA linear GSS 25-AUG-2003
LOCUS
DEFINITION
CG3AG33TV ZM_0.7_1.5_KB Zea mays genomic clone ZMMBma0754F17,
genomic survey sequence.

ACCESSION CG301209

VERSION CG301209.1

KEYWORDS GSS.

SOURCE GI:34215423

ORGANISM

Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD
Clade; Panicoideae; Andropogoneae; Zea.

REFERENCE

1 (bases 1 to 812)
Whitehead, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfling, T.,

Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.
Consortium for Maize Genomics
Unpublished (2002)
Other_GSSs: OG3AG33TH
Contact: Cathy Whitelaw
TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5943
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: methylation filtered.
Location/Qualifiers

FEATURES

source
1..812
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBma0754F17"
/note="Vector: pBCK-; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN

Query Match 5.2%; Score 97; DB 10; Length 812;
Best Local Similarity 51.2%; Pred. No. 2.6e-10;
Matches 285; Conservative 0; Mismatches 260; Indels 12; Gaps 2;
QY 427 CAGCGCGGTGGACCGGTGCTTGTACTGAGACGCGCGCGCGCATGTTCTCGCGGTC 486
DB 114 CCGCGCGCCAGCGCCACGTGCTCCGCGCGGAGCCACGCGCGCGGATGCTCGTGGCGCA 173
QY 487 TGTGTGTCATCGCTCAGGGAGCGCTCCGCTGCCAGGAGGAGGACCATCGCTCGCTGATCG 546
DB 174 TCAACACCTCTCTCCAGGGCTACTCCGCGCATCCGCTTCGAGATTTCTGAGGCCATCGCCA 233
QY 547 ACCTGTCTCAATTCCGAGCTCGCTCCGCGCGCTTCCAGCGCGCGCACGGTGGGCGGTCGG 606
DB 234 AGTGTCTCAACGCCAAGTCAAGCGCTGCTGCGCTCCGGGGCACCATCACCGCGTCGG 293
QY 607 GTGACCTGACACCGCTTGGGCATATGGTCTTGCCTTCAGGGCCGGGGAGAGATTTC--- 662
DB 294 GCGACCTCGTCCCGCTCTCTCATATCGCGGCTCTCATACGCGCGCGCCAGAACTCCGTCG 353
QY 663 --CTGGACCGGAGCGGACGCGGCTTTCAGCGCGCAGAGGGCTCCGCGCGGACGGCTGC 720
DB 354 CGGTGGGCCCGGACGAGAGGTGGCGCCCGGAGGGTTTCAGATCCGCGGATCG 413
QY 721 AACCG-----CTCGATCTCTCCCATCGCATGCACTGGCGCTGGTCAACGGGACCTCCG 774
DB 414 AGCAGGGTTCTTCGAGCTGCAGCCCAAGAGGGCTTCGCCATGGTCAACGGCACCGCGC 473
QY 775 CAGTACCGGGATCGCGCTGGTGAATGCTCACGCTTCGCCCATCTTCGGGCACTGGGCGG 834
DB 474 TGGGCTCCGCGCTTCGCTCCACCGCTCTCTTCAGGCGCAACGTTCTCGCGCTCATGGCG 533
QY 835 TGGCGTTTGAACGCGCTGCTTTCGGAATGTCAGAGGCGCGGACCGAGGCATGGCGCGCG 894
DB 534 AGGTGATCTCCGCGGTGTTCTGCGAGGTGATGACCGGCAAGCCCGAGTTTCCAGCACACC 593
QY 895 CACTGTCCGACCTTCGCGCGCATTCGCGACAGAGGACCGCGCAGGAGGCTTCGCGGCCC 954
DB 594 TGACGACAAAGCTGAAGCACCACCCCGGACAGATCGAGGCGCGCTATCATGAGGACG 653
QY 955 GCGTGGACGCGCGCG 971
DB 654 TCCTGGAAGGCGAGCTCG 670

RESULT 8

CG362861/c

LOCUS

DEFINITION

CG362861 845 bp DNA linear GSS 26-AUG-2003
OG3AZ10TH ZM_0.7_1.5_KB Zea mays genomic clone ZMMBma0766A20,


```

genomic survey sequence.
CG362861.1 GI:34280128
GSS.
SOURCE
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 845)
Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Renick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
Consortium for Maize Genomics
Unpublished (2002)
Other_GSSs: OG3AZ10TV
Contact: Cathy Whitelaw
TIGR
7712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TR
Classes: methylation filtered.
Location/Qualifiers
1..845
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBMA0766A20"
/clone_lib="ZM_0.7.1.5_KB"
/note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN
Query Match 5.2%; Score 97; DB 10; Length 845;
Best Local Similarity 51.2%; Pred. No. 2.6e-10;
Matches 285; Conservative 0; Mismatches 260; Indels 12; Gaps 2;

Qy 427 CCAGCGCGTGGGACCGGTGCTTCTGACGTGGAGCGCGCGCCATGGTTCTGGCGGCTC 486
Db 620 CCGGGCGGACGGCCACGTGTGCGGCCGAGGCGACGGCGCGCGGATGTCGTGGCGCA 561
Qy 487 TGGTGTGATCGCTCAGGAGCGCTCCGGTGCACGAGGGGACCATCGCTCGCTGTATCG 546
Db 560 TCAACACCCCTCCTCAGGGCTACTCCGGCATCCGCTTCGAGATTCTGGAGGCCATCGCCA 501
Qy 547 ACCTGCTCAATTCGAGCTCGCTCGGGCGGTTCCAGCGCGGCGACGGTGGGCGGCTCGG 606
Db 500 AGCTGTCTAACGCCAACGTCAAGCGGCTGCTGCGGCTCCGGGGACCATCAACCGGCTCGG 441
Qy 607 GTGACCTGACACCGCTTGGCATATGCTGCTCTGCTCCAGGGCGGGGAGACTTC---- 662
Db 440 GCGACCTCGTCCGCTCTCTACATCGCGGCTCATCAGGGCGCGGAGACTTCGCTCG 381
Qy 663 --CTGGACCGGGACGGGACGGGCTTGAAGCGCGCAGAGGAGGCTCCGGCGCGGACGGTGC 720
Db 380 CGGTGGGCGCCGACGGCAGGAAGGTGGGCGCGCGGAGGCTTCAGGATCGCGGCATCG 321
Qy 721 AACCG-----CTGATCTCTCCATCGCGATGACCTGGGCGCTGTCACGGGACCTCGG 774
Db 320 AGCAGGGTTCTTCGAGCTGAGCCCAAGGAGGCGCTCGCATGTTCAACGGGACCGCGG 261
Qy 775 CCATGACCGGGATCGGCTGGTGAATGCTACAGCTGCGCGCATCTCGGCAACTGGGCGG 834
Db 260 TGGGCTCGGCGCTCGCTCCACCGTGTCTTCGAGGCGCAAGTTCCTGCGGTGATGGCGG 201
Qy 835 TGGCGTTGACGGCGCTGCTTTCGGAATGTCTGAGAGGCGGACCGGAGCATGGGCGCGG 894
Db 200 AGGTCACTCCGCGGTGTTCTCGAGGTCATGACCGGCAAGCCGAGTTCACCGACCCACC 141
Qy 895 CACTGTCCGACCTCGGCGCGCATCCCGGACAGAGGAGCGCGGAGGCTGGCGCGCC 954

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 888)
Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Renick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
Consortium for Maize Genomics
Unpublished (2002)
Other_GSSs: OGNAF84TH
Contact: Cathy Whitelaw
TIGR
7712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TR
Classes: methylation filtered.
Location/Qualifiers
1..888
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBMA0325M24"
/clone_lib="ZM_0.7.1.5_KB"
/note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN
Query Match 5.2%; Score 97; DB 9; Length 888;
Best Local Similarity 51.2%; Pred. No. 2.6e-10;
Matches 285; Conservative 0; Mismatches 260; Indels 12; Gaps 2;

Qy 427 CCAGCGCGTGGGACCGGTGCTTCTGACGTGGAGCGCGCGCCATGGTTCTGGCGGCTC 486
Db 304 CCGGGCGGACGGCCACGTGTGCGGCCGAGGCGACGGCGCGGCGATGTCGTGGCGCA 363
Qy 487 TGGTGTGATCGCTCAGGAGCGCTCCGGTGCACGAGGGGACCATCGCTCGCTGTATCG 546
Db 364 TCAACACCCCTCCTCAGGGCTACTCCGGCATCCGCTTCGAGATTCTGGAGGCCATCGCCA 423
Qy 547 ACCTGCTCAATTCGAGCTCGCTCCGGCGGTTCCAGCGCGGCGACGGTGGGCGGCTCGG 606
Db 424 AGCTGTCTCAACGGCCAACGTCAAGGAGGCTGCTGCGGCTCCGGGGACCATCAACCGGCTCGG 483
Qy 607 GTGACCTGACACCGCTTGGCATATGCTGCTCTGCTCCAGGGCGGGGAGACTTC---- 662
Db 484 GCGACCTCGTCCGCTCTCTACATCCCGGCTCATCAAGGCGCGGAGACTTCGCTCG 543
Qy 663 --CTGGACCGGGACGGGACGGGCTTTCAGCGCGCAGAGGAGGCTCCGGCGCGGACGGTGC 720
Db 544 CGGTGGGCGCCGACGGCAGGAAGGTGGGCGCGCGGAGGCTTCAGGATCGCGGCATCG 603
Qy 721 AACCG-----CTGATCTCTCCATCGCGATGACCTGGGCGCTGGTCAACGGGACCTCGG 774

```

```

Db      604 AGCAGCGGTTCTTCGAGCTCGAGCCCAAGAGGGCCCTCGCCATGGTCAACGGCACCGCCG 663
Qy      775 CCATGACCGGGATCGCGCTGGTGAATGCTACAGCGCTCGCGCATCTCGGCAACTGGCGCG 834
Db      664 TGGGCTCGGCGCTCGCTCCACCGTGTCTTCGAGGCCAAGTTCTTCGCGTCAATGGCCG 723
Qy      835 TGGCGTTGACGGCGCTCTTCGCGGAATGTCTGAGAGCCCGGACCGAGGSCATGGGCGCGG 894
Db      724 AGGTCACTCTCGCGGTTCTTCGAGGTCATGACCGGCAAGCCGAGTTTACACGACCAACC 783
Qy      895 CACTGTCCGACCTCGGCGCGCATCCCGACAGAAGGACCGCGACGAGGCTGGCGCCCC 954
Db      784 TGACGCAACAAGCTGAAGCACCAACCCCGACAGATCGAGGCGCGCGCTATCATGAGACAG 843
Qy      955 GCGTGGACGCGAGCGCG 971
Db      844 TCCTGGAGGCGAGCTCG 860

```

RESULT 10
CF842933
LOCUS CF842933 737 bp mRNA linear EST 30-OCT-2003
DEFINITION pSHB023xA13f USDA-IPAFS:Expression of Phytophthora sojae genes during infection and propagation_sHB Phytophthora sojae cDNA clone sHB023A13 5, mRNA sequence.

ACCESSION CF842933
VERSION CF842933.1 GI:38058587
KEYWORDS EST.
SOURCE Phytophthora sojae
ORGANISM Phytophthora sojae

REFERENCE 1 (bases 1 to 737)
Tyler,B. Not Published
Tyler,B. Not Published
Tyler lab
VBI

1880 Pratt Dr., Blacksburg, VA 24061, USA
Tel: 540-231-7318
Email: bmytyle@vt.edu
PCR Primers
FORWARD: BK reverse primer
BACKWARD: BK reverse primer
Plate: 023 row: A column: 13
Seq primer: BK reverse primer
High quality sequence stop: 737.

FEATURES
source
1..737
/organism="Phytophthora sojae"
/mol_type="mRNA"
/db_xref="taxon:67593"
/clone="sHB023A13"
/tissue_type="mycelium"
/cell_line="p6497"
/dev_stage="48 hr. post infection stage"
/lab_host="Soybean plant"
/clone_lib="USDA-IPAFS:Expression of Phytophthora sojae genes during infection and propagation_sHB"
/notes="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"

ORIGIN
Query Match 5.1%; Score 94.4; DB 7; Length 737;
Best Local Similarity 50.0%; Pred. No. 9.3e-10;
Matches 262; Conservative 0; Mismatches 261; Indels 1; Gaps 1;
Qy 310 TCATCCGCGAGCGCGCATGTCTACGGACTGACAAACCGGCTTCGGTCCCTTGGCAACC 369
Db 213 TCAGGCACAGACGGCGGTGCGCTACGGTATCAACAGCGGCTTCGGACTCTTCCACAG 272
Qy 370 GCCTGATCTCAGGTGAGAATGTCCGAACGCTCGAGGCCAAATCTTGTCATCATCTGGCCA 429

```

Db      273 TCATCATCGCCCGGAGAGCTCACGGAGCTGAGGAGAACCTCATCCGATCGCACTCGT 332
Qy      430 GCGCGGTGGGACCGGTGCTTGAATGCTGACGACGACGCGCGCCATGCTTCTGGCGGTCTGG 489
Db      333 CGGGCACGGCGGAACCGCTGTCCCGCAGCAGACCCCGCATGTGCTGCGACTCGGCATCA 392
Qy      490 TGTGATCGCTCAGGAGAGCCTCCGCTGCCAGCGAGGGGACCATCGCTCGCCTGATCGACC 549
Db      393 ACGTGCTGGCAAGGCCCACTCAGGCATCGTGTGCATACGCTGGAGCAGCTCGTGACG 452
Qy      550 TGTCTAATTCGAGCTCGCTCCGCGCTTCCAGCGCGCGGACGCGTGGGCGCGTTCGGGTG 609
Db      453 CCTTCAACGCGGACTGCTCTGCTGGTGGTCCCAAGGAGGCACGTCCGGCGCTCGGGAG 512
Qy      610 ACCTGACACCGCTTCGCGCATATGCTGCTCTCCAGGCGCGGGGAGACTTCTCTGAGACC 669
Db      513 ATCTGGCCCCGCTCGCTCACCTCGCTTGGGATGATGGGCGAAGGCCCATGTGGGACA 572
Qy      670 GGGACGGGACGCGGCTTGACGGCGCAGAGGGGCTCCGGCGCGGACGCGCTGCAACCGCTCG 729
Db      573 AGTTGGCGAGGAGTTGCTCATCAGCG-AGGCGTCCAAGTGTGCGCAAGAACAGCCTA 631
Qy      730 ATCTCTCCATCCGATGCACTGGCGCTGTCAACGGGACCTCCGCCATGACCGGGATCG 789
Db      632 AGCCCGTGCAGCTCGGTGCCAAGGAAGTCTGGCCATGATCAACAGGCGCACGAGCTCATCA 691
Qy      790 CGCTGGTGAATGCTACGCTGCGGCATCTCGGCAACTGGGCG 833
Db      692 CGTCTGCTGGTGTGAGGCGGTGCTCCGCGTGAGAAACGTGCGG 735

```

RESULT 11
DN640249/c

LOCUS DN640249/c
DEFINITION UMC-bend 0A01-030-f10 Uterus (endometrium) bend Bos taurus CDNA 3', mRNA sequence.

ACCESSION DN640249
VERSION DN640249.1 GI:61945444
KEYWORDS EST.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.

REFERENCE 1 (bases 1 to 788)
Prather,R.S., Antoniou,E., Garverick,H.A., Green,J.A., Lucy,M.C., Roberts,R.M., Smith,M.F. and Youngquist,R.S.
USDA Grant NRI-2002-03476: Bovine ESTs: Focus on Female Reproduction

JOURNAL Unpublished (2002)
COMMENT Contact: DNA Core Facility (Bovine Project)
Animal Science - RS Prather
University of Missouri-Columbia
M616 Medical Sciences Bldg., Columbia, MO 65212, USA
Tel: (573)882-0428
Fax: (573)884-5552
Email: bovine@net.missouri.edu
POLYA=No.

FEATURES
source
1..788
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/clone_lib="bend"

/note="Funding: The production of ESTs submitted in this project was funded by USDA Grant MRI-2002-03476 entitled 'Bovine ESTs: Focus on Female Reproduction' to RS Prather, E Antoniou, HA Garverick, JA Green, MC Lucy, RM Roberts, MF Smith and RS Youngquist. Genetic Source: Heifers for the project were purchased from Circle A Ranch, Iberia, MO (http://www.circlea ranch.com/home.html). these heifers, while not registered known Angus pedigrees going back

at least 4 generations. Samples collected: The samples consisted of the following: Germinal vesicle-stage oocytes; in vitro derived embryos (2-cell, morula, blastocyst and nuclear transfer blastocyst); in vivo blastocysts and conceptuses (days 8, 14, 16 and 18); corpora lutea (days 3, 5, 8, 14, 16, 18 and 35); ovarian follicles (days 0, non-recruited, recruited, early selected and preovulatory); oviduct (days 0, 3 and 5); endometrium (days 5, 8, 14, 16, 18 and 35); and placenta/embryo from day 35 conceptuses. Expanded descriptions of how the tissues were collected can be found at the following URL:

<http://genome.rnet.missouri.edu/Bovine/Methods.html>
Library construction (Standard Protocol): All procedures have been described in detail elsewhere (Soares et al., 1994; Bonaldi et al., 1996; Jiang et al., 2001). Total cellular RNA from each sample was isolated by using STAT-60 reagent (Tel-Test, Friendswood, TX) and the poly(A)⁺ RNA was obtained by two rounds of purification with the Oligotex mRNA isolation kit (Qiagen) according to the manufacturer's instructions. The libraries were constructed by E. Ferguson and R. Woods essentially as described by the manufacturer's instructions provided with the SuperScript Plasmid System (Invitrogen, cat. no. 18248-013). Briefly, 1mcg of poly(A)⁺ RNA was annealed at c37 degrees with 10mcg of NotI-tag-dT18 oligonucleotide (GCTGTCGGCGCGC-tag-T18) and reverse transcribed at c37 degrees with Superscript II (Invitrogen) reverse transcriptase (Jiang et al., 2001). The 'tag' represents a tissue/stage-specific ten-base sequence identifier (<http://genome.uiowa.edu/pubsoft/software.html>) present in the oligonucleotide used to prime first-strand synthesis. Second strand synthesis was performed with T4 DNA polymerase in the presence of DNA ligase and RNase H. After second strand synthesis, the double-stranded cDNAs were ligated to SalI adapters (Invitrogen-Life Technologies) and digested with NotI. The cDNAs were size selected by passage through cDNA size fractionation columns (Invitrogen-Life Technologies). The cDNAs derived from each developmental stage of a particular tissue were mixed on an equimolar basis and ligated directionally into the NotI and SalI sites of the pCMV-SPORT6 vector (Invitrogen). After ligation of the inserts, the plasmids were electroporated into DH10B bacteria. Library Construction (PCR Protocol): The amount of mRNA that was recovered from oocytes and embryos was quite limiting and was not sufficient for library production with the standard protocol. Therefore, PCR-based protocol was utilized for producing libraries from sources in which the amount of extracted mRNA was small (oocytes and embryos). Poly-A RNA was isolated by using the MicroPoly(A) Pure kit from Ambion (cat. # 1918). The mRNA was reverse transcribed with a NotI-tag-dT18 oligonucleotide and a SMART oligonucleotide (Clontech) modified to contain a SalI site to generate full-length cDNA with a sequence complementary to the SMART oligonucleotide. Sequences within the SMART and dT oligonucleotides were used as primers to amplify the cDNAs by PCR with pfu turbo polymerase (Stratagene). The resulting PCR products were purified, digested with NotI and SalI and size fractionated by using Chroma Spin-1000 columns (Clontech). Purified cDNA from each PCR reaction was quantitated and mixed on an equimolar basis for ligation into the pCMV-SPORT6 vector. Preliminary Library Characterization: Randomly chosen clones from each library were analyzed by restriction digestion to determine average insert size (96 clones) and by sequencing (~4 96-well plates) to confirm library quality (e.g. the presence of short polyA⁺ tails, genomic DNA contamination (must be <1%), ribosomal RNA clones (must be <1%), etc.) and to provide a sequence database representing the predominant clones in each library. The clones were sequenced at the University of Missouri-Columbia DNA Core Facility. After production of

the libraries, equal numbers of recombinants from each library were pooled to produce a single mixed library (mega-library) for more extensive sequencing. Bioinformatics work was performed by GK Springer's bioinformatics group (Spollen WG, Topinka CM, Khambati AA) in Computer Science at the University of Missouri-Columbia. Clone Requests: Requests for clones should be made to the Director of the University of Missouri DNA Core facility at: bovine@net.missouri.edu. Bonaldi MF, Lennon G, Soares MB. Normalization and Subtraction: Two approaches to facilitate gene discovery. Genome Res. 1996; 6:791-806. Jiang H, Bivens NJ, Ries JE, Whitworth KM, Green JA, Forrester LJ, Springer GK, Didion BA, Machialagan N, Prather RS, Lucy MC (2001) Constructing cDNA libraries with fewer clones that contain long poly(da) tails. Biotechniques 31:38-42. Soares MB, MF Bonaldi, P Jelene, L Su, L Lawton, A Efstratiadis. 1994. Construction and characterization of a normalized cDNA library. Proc Natl Acad Sci, 91:9228-9232. TAG_TISSUE=Uterus (endometrium) TAG_SEQ=Not found"

ORIGIN

	Query Match	5.0%	Score 92;	DB 8;	Length 788;
	Best Local Similarity	47.8%	Pred. No. 3.1e-09;		
	Matches 300;	Conservative 0;	Mismatches 325;	Indels 3;	Gaps 1;
Qy	313	TCCCGCAGGCGCGCATGCTTACGGAGCTGACAAACGGCTTCGGTCCCTTCGGAACCGCC	372		
Db	628	TGCGGAGAACCGCACCCTACGGCATCAACCGGTTTCGGCTGTGGCCCTCGACTC	569		
Qy	373	TGATCTCAGGTGAGAAATGTGGAACGGTGCAGGCGCAATCTTGTCCATCATCTGGCCACGG	432		
Db	568	GCATCGCGCAGGAAGACCTGGAACACCTTCAGCGCTTCCTGGTGTGTCCACGCGCTG	509		
Qy	433	GCGTGGACCGGTCTTGTACTGGACAGCGCGCGCCATGCTTCTGCGCGTCTGTGGT	492		
Db	508	CGGTGGGTGAGCCCATCAGCATCGCTGTGTGGCGGTGTGTCATGGTGTCAAGGTCAACA	449		
Qy	493	CGATCGCTCAGGGAGCCCTCCGGTCCAGCGAGGGGACCATCGCTCGCTGATCGACCTGC	552		
Db	448	GCCTGAGCCGTGGTTCCTCCGGATTTCGCGGAGGTTCATCGACGCGCTGATCGCCCTGA	389		
Qy	553	TCAATTCGAGCTCGCTCGGCGCGTTCCTCAGCGCGGCAACGTTGGGCGCGTGGGTGACC	612		
Db	388	TCAACGCGGAGGTGTACCCGCACATTCGTTGAAGGTTTCGGTGGTGGTCCCTCCGCTGACT	329		
Qy	613	TGACACCGCTTGCGCATATGTGTCTCTCCCTCCAGGGCGGGGAGACTTCCCTGGACCGGG	672		
Db	328	TGGCGCCATTGGCCCAATGTCGTGCTGTGTGCTGTGGGCGAAGGCAAGGCCCGTTAC---	272		
Qy	673	ACGGGACCGCGCTTGACGGCGCAGAAAGGCTCCGGCGCGGACGCTGCAACCGCTCGATC	732		
Db	271	AAGCGAATGGCTGGGAGCACACGAGGCGCTGAAAGTTGCGGCTGACGCGCTGACCC	212		
Qy	733	TCTTCCCATCGGATGCACTGGCGCTGTGTCAACGGGAACCTCGCCATGACCGGGATCGCGC	792		
Db	211	TGGCGCCAAAGAAGGCTTGGCCCTGCTCAACGGCACTCAGGTGTCCACCGCTTATGCC	152		
Qy	793	TGGTGAATGCTCAGCCCTGCGGCCATCTCGGCAACTGGGCGGTGGCGTTGACGGCCCTGC	852		
Db	151	TGCGCGCCCTGTTTCAAGGGCGAAGACCTGTTTCGCGCGGTGCAATTGGCCCTGTGGCGCCCTGA	92		
Qy	853	TTGCGGAATGTCTGAGAGGCGGACCGAGGCAATGGGCGCGGCACTGTGCCACCTTCCCGGC	912		
Db	91	CGGTGGGAAGCCGTACTGGTTTCGGCTCGCGTTTCGAGCGCAGCATTCACGCTGCCCGT	32		
Qy	913	CGCATCCCGCAGACAGAGAGCGCCGACG	940		
Db	31	GCCAGCGTGGCCAGATCGACTCTCGCGC	4		

RESULT 12


```
Db GCATCGCCAGCGAAGACCTTGAAGAACCTTACGGTTCCCTGGTGTGTGTCCTCCACGGCGCTG 509
QY GGTGGGACCGTGTGATCGAGACGCGCGCGCGCATGTTCTGGCGGTCTGGTGT 492
Db GGTGGGAGAGCCATCAGCATGCGGTGGTGTGTCATGTTGTCAGAGTCAACA 449
QY CGATCGCTCAGGAGCGCTCCGGTGCAGAGGAGGAGCATCGCTCGCTGATCGACCTGC 552
Db GCCTGAGCGGTGTTCTCCGGGATTCGCGGACAGTATCGACCGGCTGATCGCCCTGA 389
QY TCAATTCGAGCTCGCTCCGGCGTTTCCAGACCGCGGACGAGTGGCGCGTGGGTGACC 612
Db TCAACGCGAGGTACCGGCACATTCCTGTTGAAGGTTTCGGTGGTGGCTCCGGTACT 329
QY TGACACCGCTTGGCATATGTCCTGCTCCAGGCGCGGAGAGTCTTCTGGAGCGGG 672
Db TGGCGCCATTCGCCACATGTCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 328
QY ACGGACCGCGCTTCAGCGCGCAGAAAGGCTCCGGCGGAGCGCTGCAACCGCTCGATC 732
Db AAGCGAATGCTGGAGCACCGAGGCGCTGAAAGTTGCGGCGCTGAGCGCGCTGACCC 212
QY TCTCCATCGGATGCTGCGGTGCTCAACGGGACCTCGCCATGACCGGATCGGCG 792
Db TGGCGCCAAAGAGGCTTGGCGCTGCTCAACGGGACCTCAGGTGTCACCGCTTATGCC 152
QY TGGTGAATGCTACCGCTGCGGCATCTCGGCAACTGGCGGTGGCTGTGAGCGCCCTGC 852
Db TGGCGGCGCTGTTGGAAGGCGAAGACCTGTTGCGCGGTGTCATTTGGCTGTGGCGCTGA 92
QY TTGGGGAATGCTGAGAGCGGACCGAGGATGGCGCGGCGGCTGCGACCTGCGACCTGGCG 912
Db CGGTGAAGCGGCTGCTGGGTTCGCGCTCGCGCTTCGACGACGATTCACGCTGCGCGTG 32
QY CGCATCCCGGACAGAAAGCGCGCGAGC 940
Db GCCAGCGTGGCAGATCACTCTGGCG 4
```

RESULT 13

DR785713
LOCUS ZM_BFBF_Zea_mays cdna 5', mRNA linear EST 27-JUL-2005

DEFINITION DR785713
ACCESSION DR785713
VERSION DR785713.1
KEYWORDS GI:71299412
SOURCE EST.

ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 847)
AUTHORS Kim, H., Collura, K., Wisotski, M., Smart, D., Kudrna, D., Muller, C.,
Rao, K., Haller, K., Wing, R., Soderlund, C., Walbot, V. and Yu, Y.
TITLE Maize Full-length cDNA Project
JOURNAL Unpublished (2005)
COMMENT Contact: Yeisoo Yu
Arizona Genomics Institute
The University of Arizona
Forbes Building Room 303, Tucson, AZ 85721-0036, USA
Tel: 520 626 9585
Fax: 520 621 1259
Email: yeisoo@genome.arizona.edu
Plate: 0002 row: F column: 14.

FEATURES

source
1. 847
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="B73"
/db_xref="taxon:4577"
/tissue type="mixed (silks, husks, ears, pollen, shoot
tips, leaf, root tips, whole seed, embryo)"
/dev_stage="varies by tissue"

/lab_host="DH10B T1 phage resistant"

/clone_lib="ZM_BFB"

/note="Vector: pCMV-SPORT 6.1; Site 1: EcoRV; Site 2:

NotI; Maize Full length cDNA library (3530 library)
created by Invitrogen from multiple tissues; Organ: silks,
husks, ears, pollen, shoot tips, leaf, root tips, whole
seed, embryo. This is a Gateway compatible vector, for
permitting clone movement to new vector backbones, for
expression in diverse host cells using recombination
rather than restriction enzymes. poly(A)+ mRNA was
prepared by Invitrogen, and equimolar amounts of RNA from
each of the 12 tissue samples were mixed together for
a normalization step was conducted against the mixture of
RNA samples. tissues prepared: 1. just emerging silks; 2.
inner husks from ears of sample #1; 3. 20 day aleurone; 4.
immature tassels, stages from 1-2 mm to 1-2 cm; 5. 2 mm to
2 cm ears; 6. pollen; 7. 1 cm vegetative shoot tips from
15 day old seedlings; all leaves with an expanded or
partially expanded sheath were removed; 8. mature leaf
tissue; 9. 0.5 cm long root tips from 15 day old
seedlings; 10. 10 day whole seed; 11. 12 day endosperm and
embryo; 12. 17 day endosperm and embryo. All of the
sequenced clones in Maize Full-length cDNA Project will be
archived at the University of Arizona. Clones, high
density filters and amplified library can be ordered from
the University of Arizona
(http://www.genome.arizona.edu/orders/)."

ORIGIN

Query Match 4.9%; Score 90.6; DB 8; Length 847;
Best Local Similarity 49.5%; Pred. No. 6.2e-09;
Matches 361; Conservative 0; Mismatches 344; Indels 24; Gaps 4;

```
QY 266 GGCGCGGACCGTGGCTCGGAAGCGCGGCTCGGCTGTCATCCGCGAGCGCG 325
Db 25 GGCGCGGCGCGCGCTCAAGGCGAGGAGTGTCTGATGCTGCTGCGCCACGCGG 84
QY 326 CCATGTCTACGAGTGAACAACCGGCTTCGGTCCCTTCGGAACCGCCTGATC----- 377
Db 85 CGACATCTACGGCTCAACACCGGCTTCGGCGGACCTCCACCGCGCACCAAGGACGG 144
QY 378 TCAGGTGAGATGTCGGAACGCTGCGAGCCCAATCTTGTCCATCATCT-----GGCCAGCGG 433
Db 145 GCCCGCGCTCCAGTTCGAGCTGCTCAGGCATCTCAACCGCGGAATCTTCGCGACCGGCGAG 204
QY 434 CGTGGACCGGTGCTTGACTGGACGACGCGCGCGCATGTTCTGGCGCGCTCTGGTGTGTC 493
Db 205 CGACGGGACACGCTGCGGTGGAGGTACACCGCGCGCGGATGCTGGTGGCATCAACAC 264
QY 494 GATCGCTCAGGAGCGCTCCGGTGCAGGAGGAGACCATCGCTCGCTGTGATCGACCTGCT 553
Db 265 CCTCTCCAGGGTACTCCGGCATCCGCTTCGAGATCTTCGAGGCGCATCAGAAAGCTGCT 324
QY 554 CAATTCCGAGCTCGCTCCGGCGGTTCCAGCGCGGACGCGTGGCGCGTGGGTGAGCT 613
Db 325 CAACACCGGTGTCAGCCCTGCTGCGGCGACCATCATCACCGGTGCGGCGAGCT 384
QY 614 GACACCGCTTGGCATATGGTGTCTGCTCCAGCGCGGAGACTTCTCTGG-----A 667
Db 385 GGTCCCGCTCTCTACATCGCGCGCTCATACCGGCGCGCCCCAACCGCGAGCGCGTAC 444
QY 668 CCGGGAACGGGACGCGGCTTGACGCGCAGAGAGGCTCCGGCGCGGACGCGTGCACAC--- 724
Db 445 CGTCGACGGGAGGAGGTGGACGCGCGGAGGCGTTCAGATCGCGGCGATCGAGGGCGG 504
QY 725 ---GCTCGATCTCTCCCATCGGATGACCTGGCGCTGCTCAACGGGACCTCGGCCATGAC 781
Db 505 CTTCTTCAAGCTCAACGCCCAAGGAGGCGCTCGCCATCGTCAACGGGACGCTCGTGGGCTC 564
QY 782 CGGGATCGCGTGTGTAATGCTACGCTGCGCGCATCTCGGCAACTCGGCGGTGGCGCTT 841
Db 565 CGCGCTCGCGGCGCACCGGTGATGTACGAGCCAAAGTCTCTGCGCGCTCTGTGCGAGGTCT 624
```

```

QY      842 GACGGCCCTGTTGCGGAATGTTCTAGAGCGCGGACCGAGCATGGCGCGGCACTGTC 901
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      625 GTCGCGCTTCTTCTGCGAGGTCATGACGCGGACGCGGAGTACACGACCACTGACCCA 684
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      902 CGACTTCGGCGCGCATCCCGACAGAGGACGCGCAGCGAGGCTGCGCGCCCGCGTGG 961
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      685 CAAGCTGAAGCACCAACCGGGTCCATCGAGGCGCGGCCCATCATGGAGCACATCCTTGA 744
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      962 CGCAGCGGC 970
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      745 TGGCAGCTC 753
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
CB653003
LOCUS      786 bp      mRNA      linear      EST 09-APR-2003
DEFINITION      OSJNEC03J20.f OSJNEC Oryza sativa (japonica cultivar-group) cDNA
clone OSJNEC03J20 5', mRNA sequence.
ACCESSION      CB653003
VERSION
KEYWORDS
SOURCE
ORGANISM      Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE
1 (bases 1 to 786)
Jantassuriyarat,C., Gowda,M., Haller,K., Hatfield,J., Lu,G.,
Stahlberg,E., Zhou,B., Li,H., Kim,H., Yu,Y., Dean,R.A., Wing,R.A.,
Soderlund,C. and Wang,G.L.
Large-scale identification of expressed sequence tags involved in
rice and rice blast fungus interaction
Plant Physiol. 138 (1), 105-115 (2005)
1588683
Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: rwing@genome.arizona.edu
PCR Primers
FORWARD: gta aac cga cgg cca gtc
BACKWARD: gga aac agc tat gac cat g
Plate: 03 row: J column: 20
Seq primer: gta aac cga cgg cca gtc.
Location/Qualifiers
1..786
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="OSJNEC03J20"
/tissue_type="Leaf"
/dev_stage="3 week"
/lab_host="DH10B"
/clone_lib="OSJNEC"
/notes="vector: pBluescript II KS +; Site 1: EcoRI; Site 2:
XhoI; 6 hrs after inoculation with Rice Blast (C9240-1)"

ORIGIN
Query Match      4.9%; Score 90.4; DB 6; Length 786;
Best Local Similarity 49.2%; Pred. No. 6.8e-09;
Matches 377; Conservative 0; Mismatches 356; Indels 33; Gaps 4;

QY      458 GACGGCGCGGCCCATGTTCTGGCGCGCTGTGGTGTGATCGCTCAGGAGCGCTCCGGTGC 517
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      4 GACGGTCGGGCGCTCATGTCTGTGGCGCATCAACACCTCTCTCCAGGGCTACTCCGGCAT 63
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      518 CAGCGAGGGACCATCGCTCGCTTGATGACCTGCTCAATTCGAGTCTCGTCCGGCGGT 577
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db      64 CCGGTTTCGAGATCCTCGAGGCCATCAACAAGCTGCTCAACACCGGCGTCA CGCGTGCCT 123
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      578 TCCAGCCGGCGCACGTCGGCGCGCTGCGGTGACCTGACACCGCTTCGCGCATATGTCCT 637
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      124 GCGCTCCGTGGCACCATCAACCGGTCGCGTGACCTGGTTCCCTCTCTACATTTGCGGG 183
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      638 CTCCTCTCAAGGGCGGGGAGACTTCTCTGGAC-----CGGGACGGGACGCGGCTTGACGG 691
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      184 CCTCATCACCGCGCGCCCAACCGCAGGCGCATCTCGCCCGCAGCGCAGGAAGGTGGACGC 243
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      692 CGCAGAAGGCTCCGGCGCGGACGGGTGCA-----ACCGCTGATCTCTCCATCCGCA 745
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      244 CGCCGAGGCGTTCAAGCTCGCGGCATCGAGGGTGGCTTCTTCACGTGAACCCCAAGGA 303
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      746 TGCACCTGGGCTGTCAAACGGGACCTCCGCATGACCGGATCGCGCTGGTGAATGCTCA 805
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      304 AGGTCTCGCCATCGTAATGGACGTCCTGGGGGTGCGCGCTCGCGGCCACCGTGATGTT 363
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      806 CGCTTCGCGCATCTCGGCAACTGGGCGGTGGCGGTTGACGGGCCCTGCTTGGCGGAATGCT 865
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      364 CGACGCCAATCCTCGCGCTCTGTCCGAGGTGCTCTCGCGGTGTTCTCGCAGGTGAT 423
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      866 GAGAGCGCGACCGAGGATGGCGCGGCATGTGTCGACCTGCGGCGCGCATATCCCGACA 925
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      424 GAACGGCAAGCGGAGGTACACCGACCACTGACCCCAAGCTGAAGCACCACCCCTGGGT 483
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      926 GAAGGACCGCGACGAGGCTGCGCGCCGCTGGAGCGGCGAGCGCGGCTGGTCCGCA 985
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      484 GATCGAGCGCGCGCCATCATGAGAGCATCTCTCGCGGAGTCTGTTCAATGAGCCACGC 543
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      986 CGTCATTTCGCGAGCGGAGGCTCGACCGCGCGATATCGGACCGAGCGGAGCGGGCGCA 1045
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      544 CAAGA-----AGGTGAACGAGATGGACCCGCTGCTGAAGCCGAAGCA 585
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      1046 GGATGCTACAGCTGCGCTGCGCTCGGAGTTCTCGGGCGGGCTTCACACGCTCGC 1105
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      586 GGACAGGTACGCGCTCGCACGTCGCGCAGTGGCTCGGCGCCGACATCGAGTGCATCCG 645
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      1106 ATGCGATGACCGGCTGCTGACGATCGAGCTGAACGGGTGACGACAAATCCGGTGTTC 1165
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      646 CGCGGCCACCAAGTCCATCGAGCGGAGGTCACTCGGTGAACGACACCGGTGATCGA 705
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      1166 GCCCGATGGCAGGTGCCCGCTTCGACCGGGGCAATTTTCATGGC 1211
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      706 CGTCCACCGCGCAAG--GCGTCCACGCGCGCACTTCAGGGC 748
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
CB642693
LOCUS      866 bp      mRNA      linear      EST 08-APR-2003
DEFINITION      OSJNEB03A03.f OSJNEB Oryza sativa (japonica cultivar-group) cDNA
clone OSJNEB03A03 5', mRNA sequence.
ACCESSION      CB642693
VERSION
KEYWORDS
SOURCE
ORGANISM      Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
Jantassuriyarat,C., Gowda,M., Haller,K., Hatfield,J., Lu,G.,
Stahlberg,E., Zhou,B., Li,H., Kim,H., Yu,Y., Dean,R.A., Wing,R.A.,
Soderlund,C. and Wang,G.L.
Large-scale identification of expressed sequence tags involved in
rice and rice blast fungus interaction
Plant Physiol. 138 (1), 105-115 (2005)
1588683
Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088, USA

```

Tel: 520 626 3967
Fax: 520 621 9288
Email: rwing@genome.arizona.edu
PCR Primers
FORWARD: gta aaa cga cgg cca gtg
BACKWARD: gga aac agc tat gac cat g
Plate: 03 row: A column: 03
Seq primer: gta aaa cga cgg cca gtg.
Location/Qualifiers
1..866
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="OSJNEB03A03"
/issue_type="Leaf"
/dev_stage="3 week"
/lab_host="DH10B"
/clone_lib="OSJNEB"
/note="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
XhoI; 24 hrs after inoculation with Rice Blast (Che
86061)"

Db 696 GGCTCGGCCCGCAGATCGAGTCAATCCGCGCGGCCACCAAGTCCATCGAGCGCGAGGTCA 755
Qy 1138 ACGCGGTGACCGACAATCCGGTGTTCGCGCCGATGGCAGCGTGCCTCCCTGCAAGGG 1197
Db 756 ACTCGGTGAACGACAACCGGTGATCGACGTCCACCGCGCAAG--GGCTCCACGGCG 812
Qy 1198 GCAATTTTCATGGCCAGC 1215
Db 813 CAACTTCCAGGGGACC 830

Search completed: December 10, 2005, 17:16:23
Job time : 7544 secs

ORIGIN		Query Match	Best Local Similarity	4.9%	Score 90.4	DB 6	Length 866
		Matches 389	Conservative 0	Mismatches 376	Indels 33	Gaps 4	
Qy	430	GCGCGGTGGACCGGTGCTGACATGGACGACGGCGCGCCATGGTTCCTGGCGGTCTGG 489					
Db	54	GCTCGATGGCCACACGCTCGCGTCGGAGACGGTGGGGGGCCATGCTCGTGGCATCA 113					
Qy	490	TGTCGATCGCTCAGGAGCCTCCGGTGCACGAGGGGACCATGCTCGCTGATCGACC 549					
Db	114	ACACCTCTCTCAGGGTACTCCGGCATCCGGTTCGAGATCTCGAGGCCATACCAAGC 173					
Qy	550	TGCTCAATTCGAGCTCGCTCCGCGCTTCCAGCGCGGACGCTGGCGCGTCCGGTG 609					
Db	174	TGCTCAACACCGCGGTACGCGCTGCTCGCGCTCCGCTGGGACCATCACCGCTCCGGTG 233					
Qy	610	ACTTGACACCGCTTTCGGCATATGGTGTCTGCTCTCCAGGGCGGGGAGACTTCTTGAC- 668					
Db	234	ACCTGGTTCCCTGTCTTACATTGCCGCTCATCACCGCGCGCCCAACCGCGCAGGCCA 293					
Qy	669	-----CGGAGCGGACGGGCTTGAACGGCGAGAAAGGCTCCGGCGGGAGCGGCTGCA-- 721					
Db	294	TCTCGCCGACCGGAGGAAGTGGACGCGCGGAGGGGTTCAAGCTCGCGGCATCGAGG 353					
Qy	722	----ACCGCTCGATCTCTCCCATCGCATGCACTGGCGCTGGTCAACGGGACCTCCGCCA 777					
Db	354	GTGGCTTCTTCAAGTGAACCCCAAGAAAGGTCTCGCCATCGTCAATGGCACGTCCTGG 413					
Qy	778	TGACCGGGATCGCGTGGTGAATGCTCAGCGCTCGCGCATCTCGGAACTGGGCGGTGG 837					
Db	414	GGTCGGCGCTCGCGGCCACCGTGTATGTTGACGCGCAACATCTCGCGTCTGTCCGAGG 473					
Qy	838	CGTTGACGGCCCTGCTTTCGGAATGTCTGAGAGCCCGGACCGAGGCATGGGCCCGGCAC 897					
Db	474	TGCTCTCGCGGCTTCTCGAGGTGATGAACGGCAAGCGGAGTACACCGACACACCTGA 533					
Qy	898	TGTCGACCTCGCGCGGCTCCCGGACAGAGGAGCGGCGAGCGGCTGGCGCCCGCG 957					
Db	534	CCCAAGAGCTGAAGCACACCCCTGGGTGATTCGAGGGCGCGCCATCATGGAGCACATCC 593					
Qy	958	TGACGCGCAGCGCGGCTGGTCCGGCACGTCATTGCGAGCGGAGGCTCGACCGCGCG 1017					
Db	594	TCGCGGGAGCTCGTTTCATGAGCCACGCCAAGA-----AGGTGAACG 635					
Qy	1018	ATATCGGGAACGAGCGGAGCGGGGAGGATGCTACAGCTCGCTGCGCTCCGAGG 1077					
Db	636	AGATGGAACCGCTGCTGAAGCCGAAGCAGGACAGGTACGCGCTCCGCAACGTCGCGCAGT 695					
Qy	1078	TTCTCGGGGGGGCTTTCGACACGCTTCGCATGGCATGACCGGGTCTGACGATCGAGCTGA 1137					

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 10, 2005, 02:01:09 ; Search time 343 Seconds
(without alignments)
9587.432 Million cell updates/sec

Title: US-10-621-826-2

Perfect score: 1850

Sequence: 1 ccagcgccatccacgag.....tcagctcgatcccgagagg 1850

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents NA.*
- 1: /cgn2_6/ptodata/1/ina/1 COMB.seq.*
 - 2: /cgn2_6/ptodata/1/ina/5 COMB.seq.*
 - 3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
 - 4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
 - 5: /cgn2_6/ptodata/1/ina/H COMB.seq.*
 - 6: /cgn2_6/ptodata/1/ina/PCUS COMB.seq.*
 - 7: /cgn2_6/ptodata/1/ina/PP COMB.seq.*
 - 8: /cgn2_6/ptodata/1/ina/RE COMB.seq.*
 - 9: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	215.4	11.6	1527	3	US-09-902-540-8003
2	215.4	11.6	6439	3	US-09-902-540-813
3	171.8	9.3	1548	3	US-09-252-991A-16475
4	171.8	9.3	2295	3	US-09-252-991A-16126
5	163.2	8.8	1527	3	US-09-489-039A-6404
6	161.8	8.7	1542	3	US-09-833-745-7
7	153.8	8.3	1536	3	US-09-833-745-12
8	125.4	6.8	1557	3	US-09-252-991A-16124
9	125.4	6.8	1563	3	US-09-252-991A-16478
10	110.4	6.0	2787	3	US-09-624-693A-16
11	110	5.9	963	3	US-09-252-991A-16368
12	106.6	5.8	2475	3	US-09-624-693A-20
13	103.2	5.6	2061	3	US-09-765-873A-31
14	103.2	5.6	2151	3	US-09-627-216A-7
15	103.2	5.6	2151	3	US-09-765-873A-9
16	101.6	5.5	2151	3	US-09-627-216A-9
17	101.6	5.5	2151	3	US-09-765-873A-9
18	101.6	5.5	2439	3	US-09-624-693A-18
19	100.6	5.4	2476	3	US-09-221-017B-876
20	88.4	4.8	798	3	US-09-252-991A-16232
21	83	4.5	2419	3	US-09-624-693A-12
22	78.4	4.2	3139	3	US-09-949-016-303
23	78.4	4.2	3786	3	US-09-919-039-351
24	76.8	4.2	3141	3	US-09-949-016-4130

25	67.6	3.7	2465	3	US-10-439-479-24	Sequence 24, Appl
26	66.2	3.6	438	3	US-09-252-991A-16256	Sequence 16256, A
27	63.8	3.4	1455	3	US-09-615-192A-245	Sequence 245, App
28	57.8	3.1	1320	3	US-09-724-797-85	Sequence 85, Appl
c 29	57.4	3.1	471	3	US-09-252-991A-16230	Sequence 16230, A
30	57.2	3.1	29272	3	US-09-902-540-1217	Sequence 1217, Ap
31	54.6	3.0	94750	3	US-09-596-002-38	Sequence 38, Appl
32	54.4	2.9	1512	3	US-09-902-540-4044	Sequence 4044, Ap
c 33	54.4	2.9	24986	3	US-09-902-540-1200	Sequence 1200, Ap
c 34	54	2.9	3354	3	US-09-902-540-9590	Sequence 9590, Ap
c 35	54	2.9	13346	3	US-09-902-540-1089	Sequence 1089, Ap
36	53.6	2.9	77536	3	US-09-410-551B-1	Sequence 1, Appli
37	53.6	2.9	77536	3	US-09-940-316B-1	Sequence 1, Appli
38	53.4	2.9	1500	3	US-09-252-991A-6254	Sequence 6254, Ap
c 39	53.4	2.9	1695	3	US-09-252-991A-6044	Sequence 6044, Ap
c 40	53	2.9	2561	3	US-09-616-289-48	Sequence 48, Appl
c 41	53	2.9	2561	3	US-09-976-740-48	Sequence 48, Appl
c 42	53	2.9	6360	3	US-10-212-962-2	Sequence 2, Appli
43	52.4	2.8	44377	2	US-08-804-227C-7	Sequence 7, Appli
44	52.4	2.8	44377	2	US-08-804-198-1	Sequence 1, Appli
45	51.8	2.8	774	3	US-09-252-991A-15298	Sequence 15298, A

ALIGNMENTS

RESULT 1
US-09-902-540-8003
; Sequence 8003, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 8003
; LENGTH: 1527
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-8003

Query Match		11.6%	Score 215.4; DB 3; Length 1527;
Best Local Similarity		50.0%	Pred. No. 8e-34;
Matches 753; Conservative		0;	Mismatches 691; Indels 63; Gaps 6;
Qy	203	CATCGATCTGGACCGAGCCATCGCTGGCGAGCGCGCGGATTTGCTTGTGCCCC	262
Db	36	CTTGAAGCTCGAGGAATCTCCAGTGGTTCGCAACGAGGCCACCTGGAGCTGTGCC	95
Qy	263	TCGCGCGCGAGCGGTCGCTCCGAGCGGCTCGGCGCTGTCTATCCGCGAGGC	322
Db	96	CGAGCGCGCACCGCGCTCGCGCGCTCGGCGCGCTCGGCGCGCGCGGAGA	155
Qy	323	CGCGCATGTCTACGAGTCAACCGGCTTCGGTCCCTTCGCAACCGCTGATCTCAGG	382
Db	156	CACGCCCGCTTCGCGCATCAACCGGCTTCGGCGAGTTCGCGCATCGACAA	215
Qy	383	TGAGATGTCGAGACGCTGCAGGCAATCTGTCCATCATCTGCGCAGCGCGGAGCC	442
Db	216	GAGGAGCTGCGGACCTCCAGCGCACTATCTCCACGCTGTGCGCTGCGCAC	275
Qy	443	GGTGTCTTACTGAGCAGCGCGCGCCATGTTTCTCGCGCGTGTGTTGTCATCGTCA	502
Db	276	GCCCTCCCTCCCGGAGCGGCGCTCTCTGCTGCTCGCTGCAACGCTGTGCGCAA	335
Qy	503	GGGAGCCTCGGTGCCAGCGAGGGGACCATGCTCGGCTGATCGACTGCTCAATTCCGA	562

```
Db 336 GGGCTACTCCGGCATCCGCATGGAGACGCTGGCCCTGGGCATGCTGAACCGGA 395
Qy 563 GCTCGCTCCGGCCGTTCCAGCGCGGACGCTGGGCGCGTTCGGTGAACCTGACACCGCT 622
Db 396 CGTGTGCGCGGTGCTCCCGAGCGGGGACGCTGGGGCGCTCCGGGATCTCGCCCGCT 455
Qy 623 TGGGCATATGCTGCTGCTCCAGGCGCCGGGAGACTTCTGACCGGACCGGACGCG 682
Db 456 GCGCACCTTGGCTGCTCTTCATCGCGAAGGTGAAGCCTTCTATCAGGGCCA--GCG 512
Qy 683 GCTTGACGGCGCAGAAGGGTTCGGGCGCGGACCGGCTGCAACCGCTCGATCTTCCCATCG 742
Db 513 GATGCCCGCAGACGCGCTGGAGCGCGCGCGCTGCAACCGGTGTGTGGAGGCCAA 572
Qy 743 CGATGCACTGGCGTGGTCAACGGGACCTCGGCCATGACCGGATCGCGTGGTGAATGC 802
Db 573 GGAGGGGCTCGCCCTGGTGAACGGGACACAGGCCATGTGCGCGTGGGACCCCTGTCCA 632
Qy 803 TCACGCGCTGCGGCATCTCGGCAACTGGGCGGTGGCGTTGACGGCCCTGCTTGGGAATG 862
Db 633 GCTTCGGCGGAGTCCCTGGCGGACATCGCGGAGCTCGGGGCGCATGACGCTGGAGG 692
Qy 863 TCTGAGAGCGCGACCGAGGATGGGCGCGGCACTGTTCGACCTGGCGCGCATCCCGG 922
Db 693 GCTGCTGGGAAGCCACAAGCCCTTCACTTCCTGAGATTACGACGTCCGCGCGCACCGGG 752
Qy 923 ACAGAGGAGCGCGCAGGAGGTGCGCGCGCGGTGGAGCGGCGCGGTGGTCCG 982
Db 753 CCAGAAGGAGCTCGCGCGCACCTGGCGGCGCATCTCTGGTGGACAGCGAGCTGGTGAATC 812
Qy 983 GCACGTCAATTGCGGAGCGGAGGCTCGACGCGCGCGATATCGGACGAGCGGAGCGGG 1042
Db 813 GCAGTCA-----ACTGCAGCAAGGT 833
Qy 1043 GCAGGATGCTACAGCTGCGGTGCGCTCCGAGGTTCTCGGGCGGGCTTCGACACGCT 1102
Db 834 GCAGGACCCCTACTCTCCCTGCGCTGCATGCGCGAGGTGACGGCGCGCGCGGAGGCGAT 893
Qy 1103 CGCATGGCATGACCGGGTGTGACGATCGAGCTGACGGGTGACGGGTGACGACATCCGGTGT 1162
Db 894 CGCGTTCTCCCGCGCATCTCGAGGTGGAGGTCAACAGCGCAGCGACAAACCCGCTCGT 953
Qy 1163 TCCGCCCGATGGCAGCGGTGCGCGCTGACGCGGGGCAATTTATGGGCGAGCATGTGC 1222
Db 954 ---GTTTCGCGGACACGAGGCGCATGTGTTCGGGCGGCACTTCCAGGCGAGCCCATCTC 1010
Qy 1223 GCTGACGCTCGATGCGCTCGCCACGCGCGTCA CGGTTCTGGCGGGCTTTCGGAGCGCCA 1282
Db 1011 CCTGGCCATGGACGCTGGTGGCGATGGCGCTGACGCAACTGTCTCCATCAGCGAGCGCG 1070
Qy 1283 GATTGCACTGACAGATGAAGGCTGAACCGTGGGCTGCCCCCTTCTCCACCGGG 1342
Db 1071 CGTGGAGCAGCTCGTGAACCCGCTCGCTGTCCAACTTGC CGCGGT-----TCTCGCGAA 1124
Qy 1343 CCCCGCGCGGTGAATTCGCGCTTCATGGGCGCACAGGTGACGCGACGCGCTCTCTGGC 1402
Db 1125 GAATCTCGGGTTGAATCTCCGGCTTCATGATCGGCGAGGTGAACAGCGCGCGCTGGTGGC 1184
Qy 1403 CGAGATGCGAGCCACGCGGA---CCTGCTCGATCCATTCGATCTCAACGAAACCGCCCAA 1459
Db 1185 CGAGTCCCGGCTGCTGAGCCACCCCGCTCCGTGGATTGATTCGTCATCCCGCGGCGC 1244
Qy 1460 TCAGGATGTTGTTCTCGTTGGGACCATCGCGCGCGCTCTGCGCGAGAGATCGACCG 1519
Db 1245 AGAGGCCACAGCTGTCATGGGCGATGACGCGCGCGCTCAAGGGCGCTCAGGTACGCACTT 1304
Qy 1520 TTGGGCGGAGATCCTTGGCAGTCTCTGCTCTGTCTTGACAAAGCTGCGGAGCTGCGCTG 1579
Db 1305 CGCCCGTTCTGCTCGCGATTGAATCTCTGTTGGCGGCGCGAGCCCTTGAATTCGCGCT 1364
Qy 1580 CGGACGCGGCTTAGACGGGGTGTCTCCCGCGGGGAAGAAGCTGGTGAGGCCCTTCGCGGA 1639
```

```
Db 1365 GCGCTGAAGCCCGGCAAGGGCGCCCTCGCGCGGTACGAGCTGG-----TGCCTC 1415
Qy 1640 GCAGTTCCCGCGCTTTGAGACGACCGGCCCTTGGGACAGAAATTCGCGCGCTTGCTAC 1699
Db 1416 GAAGGTCCCCACATGACAGGACCGCGAGCTGACCCGGGACATCGAGCGGTGAGCCA 1475
Qy 1700 GCACCTC 1706
Db 1476 GCTCGTC 1482

RESULT 2
US-09-902-540-813
; Sequence 813, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 813
; LENGTH: 6439
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-813
```

```
Query Match 11.6%; Score 215.4; DB 3; Length 6439;
Best Local Similarity 50.0%; Pred. No. 9,2e-34;
Matches 753; Conservative 0; Mismatches 691; Indels 63; Gaps 6;
```

```
Qy 203 CATCGATCTGGACACGAGCCCATGCGCTGGCGAGCGCGCGCGGATTTCTCTTGGCCC 262
Db 1355 CCTGAAGCTCGAGGAATCTCCAGGTGGCTCGCAAGAGGCCACCGTGGAGCTGGCC 1414
Qy 263 TCGCGCGCGAGACCGGTGCGGTGCGTCCGAGACGCGCGCTGCGCGTGTCTATCCGAGGC 322
Db 1415 CQACGCGCGCACCGCGTGGCGCTCGCGCGCTTGGTGAACCGCTCGCCCGCGAGA 1474
Qy 323 GCGCATGCTAGGACTGACACCGGCTTCGCTCCCTTGGCAACCGCTGATCTCAGG 382
Db 1475 CAGCCCGCTACGGCATCAACACCGGCTTGGCACTTGGCGGAGTCCGCATCGACAA 1534
Qy 383 TGAGAAATGTCGAAACGCTGCAGGCCAATCTTGTCCATCATCTGGCCAGCGCGCTGGGACC 442
Db 1535 GAAGGACCTGGCGACCTCCAGGGCAACCTCATCTCTCCAGCGCTGTGGCGTGGCAC 1594
Qy 443 GTGTCTTGAATGGAACGCGCGCGCCATGTTCTGGCGCGTCTGGTGTGATTCGATCA 502
Db 1595 GCGCCCTCCCTTCCGAAAGCGCGCGCTCTCTGCTGCTCCGCTGCAACGCTGTGCGCAA 1654
Qy 503 GGGAGCTCGGTGCGCAGCGAGGGGACCATCGCTCCGCTGATCGACCTGCTCAATTCGA 562
Db 1655 GGGCTACTCGGCATCCGATGAGAGACGCTGGGCCCTTGGCGCTGGACATGCTGAACCGGGA 1714
Qy 563 GCTCGCTCGCGCGTTCGACGCGCGGCGAGCGGTGGCGGTGCGGTGACCTGACACCGCT 622
Db 1715 GCTGTGCGCGTGTCTCCGAGCGGGGAGCGGTGGCGCTCGGGGATCTCGCCCGCT 1774
Qy 623 TCGCATATGTTGCTGTGCTCCAGGGCGGGGAGACTTCTCTGGACCGGAGCGGAGCGG 682
Db 1775 GGGCGACCTGGCGCTGCTTCTATCGGCGAAGGTGAAGCCTTCTATCAGGGCCA--GCG 1831
Qy 683 GCTTGAAGCGGACAGAGGCTCCGGCGCGGACGGCTGCAACCGCTGATCTCTCCCATCG 742
Db 1832 GATGCCCGGAGACGAGCGCTGGAGCGCGCGCGCTGCAACCGGCTGTGTGGAGGCCAA 1891
```



```
QY 1017 GATATCGGACGAGCGCGAGCGGGCGAGATGCTTACAGCCTGGCGTGGCTCCGCGAG 1076
Db 901 ---TCCCATGAGAAGTGCACAGAGTCCAGAACCTTATTCGTGTGGTTGCCAGCGCGAG 957
QY 1077 GTTCTCGGGCGGGCTTCGACACGCTCGCATGACCGGTGCTGACGATCGAGCTG 1136
Db 958 GTGATGGGCGCTGCTGACCCAGATGCGCCAGCGCGCGAGGTGCTGGAGATCGAAGCC 1017
QY 1137 AAGCGGTGACACACATCCGGTGTTCGCCCGATGGCAGCGTGGCCCTTCGACGGG 1196
Db 1018 AACGCGGTGTCCACAAACCGCTGGTATTCGCCCGCAGAGGGCG---ACGTGATCTCCGGC 1074
QY 1197 GGCAATTTCAATGGCCAGACATGTGGCGCTGACGCTCCGATGCGCTCGCCACGCGCGTCACC 1256
Db 1075 GGCAATTTCAACGCCGACCGGTGGCGATGGCCGCGCAGCAACTTGGCCCTGGGGTTGGCC 1134
QY 1257 GTTCTCGGGCGCTTCGCGAGCGCCAGATGTCACGCTGACAGATGAAGGCTGAACCGT 1316
Db 1135 GAGATCGGTTCGCTGCGAAACGGCGCATCTCGCTGATGATGA-----CATGCACATG 1188
QY 1317 GGCTCGCCCTTCCTCCACGGGGCCCCCGCGGTGATTCGCGCTTCATGGGGGCA 1376
Db 1189 TCGCATGTGCGCGCTTCCTGCTGSGCCAAACGGCGGGGTCAACTCCGGCTTCATGATCGCC 1248
QY 1377 CAGGTGACGGCGACCGCGCTCTCGCGCGAGATCGGAGCCACGGGACCTGCTCTCGATCCAT 1436
Db 1249 CAGGTACCGCGCGCGCTGGCCAGGACACAAAGGCCCTGSCCATCCGGCCAGCGTC 1308
QY 1437 TCGATTCACACGAACCGCCCAATCAGGATG 1467
Db 1309 GACAGCGTGGCGACCTCGGCCAACACGGAAG 1339
```

RESULT 4

```
US-09-252-991A-16126/c
; Sequence 16126, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16126
; LENGTH: 2295
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16126
```

```
Query Match 9.3%; Score 171.8; DB 3; Length 2295;
Best Local Similarity 49.2%; Pred.No. 3.8e-25;
Matches 606; Conservative 0; Mismatches 577; Indels 48; Gaps 4;

QY 237 GCGCGGCGCGGATGTCTTCCCTTCGCGCGCGGACCGGTGCGGTCCGAGCG 296
Db 1362 GCCCGCTGCGCTGAGCCTGGATCCAGCGCGCGCGCTGATCGCGCCAGCGTCGCC 1303
QY 297 CGGCTCGGCGCTGTATTCGCGAGGCGCGCATGTCTACGGAATGACAAACGGCTTCGGT 356
Db 1302 TGTGTGAGAAATCATCTCGCGAGGCGCGCACCGCTTACGGCATCAACACCGCTTCGCC 1243
QY 357 CCCCTTTCGAAACCGCTGATCTCAGGTGAGATGTCCGAACGCTGCGAGGCCAATCTTGT 416
Db 1242 CTGCTGGCTCGAGCGGATTTCCCGCGCGAGCTTGAGAGAGCTTCAGGTTCCATCGTC 1183
QY 417 CATCATCTGGCCAGCGCGGTGGGACCGGTGCTTGAATGGACGACGCGCGCGCATGGTT 476
```

RESULT 5

```
Db 1182 CTGTCCCATCCGCGCGCTCGCGAGGCGCTGGACGACGCCATGCTGGCGCTGGTCATG 1123
QY 477 CTGGCGGCTGTGGTGTGATCGCTCAGGGAGCCTCCGGTGCAGCGAGGGGACCATCGCT 536
Db 1122 CTGCTCAAGGTGAACAGCCTGGCGCGCGGTTCTCCGGCATCCGCGCAAGGTGATCGAC 1063
QY 537 CGCTCATGACACTGTCTCAATTCGAGCTCGCTCCGGCGGTTCCAGCGCGCGGACGCGTG 596
Db 1062 GCGCTGATCGGCTGATCAACGCGGAGGTCTATCCGCACATCCCGCTGAAAGGCTCGGTG 1003
QY 597 GCGCGCTCGGGTGAACCTGACACGCTTTCGCTATATGTTGCTCTTCAGGGCCGCGGGA 656
Db 1002 GGTGCGCTCGCGGACCTGGCGCGCTGGCGCACATGTGCT---GGTGTGATCGCGCA 946
QY 657 GACTTCTGGAACGGGACGGGACCGGCTTGACGGCGCAGAGGGCTCCGGCGCGGACGG 716
Db 945 AGCGGGCGCGCATCGCGGTGAGTGGCTGCGCGCGCGCAAGCGCTGGCGGTGGCGCGG 886
QY 717 CTGCAACCGCTCGATCTCTCCATCGCATGCACTGGCGCTGGTCAACGGGACCTTCGCC 776
Db 885 CTGGAGCGCTGACCTTGGCGGGAAGGCTGGCCCTGCTCAATGGCACCCAGGTG 826
QY 777 ATGACCGGATCGCGCTGTGTAATGTCAAGCTGCGGCCATCTCGGCAACTGGGCGGTG 836
Db 825 TCCACCGCTACGCGTGGCGGGTGTTCGAGGCGGAGGACCTGTTGCGCGCGCGCAC 766
QY 837 GCGTTAGCGCCCTGTTGGGNAATGTCTGAGAGGCGCGACCGAGGACATGGCGCGGCA 896
Db 765 GTCTGCGGGGCTCAGCGTGAAGCCATGTCTGCTGCGGGCGCGCTTCGATGCGCGC 706
QY 897 CTGTCCGACTCGCGCGCGCATCCCGGACAGAGGACGCGCAGCAGAGGTGCGCGCGCCG 956
Db 705 ATCCATCGCGCGCGCGCAGCGTGGGCAAGATCGACGTGCGCGCGCTATCGGACCTG 646
QY 957 GTGAGCGGACGCGCGGTGCTCGGACGCTGCTGCGAGCGGAGGTTCGACGCGCGGC 1016
Db 645 CTACCGCGCAGCAGGCTGGCGCGC----- 619
QY 1017 GATATCGGACGCGCGGAGGCGGGCAGGATGCTTACAGCTGCGCTGCGCTCCGCGAG 1076
Db 618 ---TCCCATGAGAAGTGCACAAAGTTCAGGACCCCTATTCGTGCTGTGCGCAGCGCGAG 562
QY 1077 GTTCTCGGGCGGGCTTCGACACGCTCGCATGGCATGACCGGGTGTGACGATCGAGCTG 1136
Db 561 GTGATGGCGCTGCTGCTGACCCAGATGCGCCAGGCGCGGAGGTGCTGGAGATCGAAGCC 502
QY 1137 AACGCGGTGACGACAACTCGGTGTTTCGCGCGATGGCAGCGTGCCTGCGCTGACGGG 1196
Db 501 AACGCGGTGTCGACAAACCGCTGTGTTTCGCGCGCGAGGGCG---ACGTGATCTCCGGC 445
QY 1197 GGCAATTTCAATGGCGCAGCATGTGGCGCTGAGCTCGATGCGCTCGCCACGGCGCTCAC 1256
Db 444 GGCAATTTCAAGCCGAAACCGGTGGCGATGGCGCGACAACTGCGCTTGGGCTTGGCC 385
QY 1257 GTTCTGGCGGGCTTCGCGAGCGCCAGATTGACGCTTTCAGACAGATGAAGGCTGAACCGT 1316
Db 384 GAGATCGGTTCGCTGTGCGAACGGCGCATCTCGCTGATGATGA-----CATGCACATG 331
QY 1317 GGGCTGCCCCCTTCCTCCACCGGGGCCCCCGGGTGTGAATTCGGGCTTCATGGGGCA 1376
Db 330 TCGCAGTTTCGCGCGTTCCTGTGGTGGCAACGGCGGGTCAACTCCGGCTTCATGATCGCC 271
QY 1377 CAGGTGACCGGACCGCGCTCTTGGCCGAGATGCGAGCCACAGGGCTTCGCTCGATCCAT 1436
Db 270 CAGGTCAACCGCGCGCGCTTGGCCAGGACAAACAGGCCCTTGGCCCATCGGCCAGGTC 211
QY 1437 TCGATCTCCACGAACCGCGCCCAATCAGGATG 1467
Db 210 GACAGCTCGCGACCTCGGCCAACACGGAAG 180
```

US-09-489-039A-6404
; Sequence 6404, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489, 039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 6404
; LENGTH: 1527
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-6404

Query Match 8.8%; Score 163.2; DB 3; Length 1527;
Best Local Similarity 47.6%; Pred. No. 1.8e-23;
Matches 655; Conservative 0; Mismatches 673; Indels 48; Gaps 4;

Qy	210	CTGACACAGGCCCATGCGTGGCGAGCGCGCGCGGATGTGCTTGCCTCCCGGCG	269
Db	37	CTGTGCGAGCTCCGCGAGGTTTACAGCCACCCGCTGAACATTAACCTTGTAGCAGCGCG	96
Qy	270	CGCGACCGGTGCGGTGCTCCGAAGCGCGGTGCGGCGCTGTATCCGCGAGCGCGCCAT	329
Db	97	TTTGGCGCAATTGATGAGAGCTGGCTGCTCAAGCCATCTCTCGCGAAGGCGGACC	156
Qy	330	GTCTACGGAAGTACAAACCGGCTTCGGTCCCTTCGGAACCGCTGATCTCAGGTGAGAA	389
Db	157	GCCTACGGCATCAACACCGGTTTGGCCCTGCTGGCGCAGACCCGTATCTCGACGGAAG	216
Qy	390	GTCGAACGCTGAGAGCCAAATCTTGTTCATCATCTGGCCAGCGGTGGAGCCGGTGTT	449
Db	217	CTGGAGAACTCTGACAGCGTTCTGCTGTGCTATCCACGCGCGGGGCTTGGCGAGCGCTG	276
Qy	450	GACTGGACGACGCGCGCGCCATGTTCTGGCGGTCTGTGTGTCGATCGCTCAGGAGGCC	509
Db	277	GACATGATCTGGCGCGGCTGATTATGGTGTGAAGATTAAACAGCTGTCCCGCGTTTC	336
Qy	510	TCCGGTGCCACGAGGGGACCATCGCTCGCTGATCGACCTGCTCAATTCGAGCTCGCT	569
Db	337	TCCGGGATCCGCTGAGCGTGATCCAGCGCTGATTGGCTTGGTCAATGCCGTGACG	396
Qy	570	CGGCGGTTCCAGCGCGGACCGTGGGCGCGTGGGTGACCTGACACCGCTTGGCAT	629
Db	397	CCGTGGATCCCGGCAAAAGGCTCCGTGGCGCTCCGCGACCTCGCGCGCTGGCGCAT	456
Qy	630	ATGTGCTGCTGCCCTCCAGGCGCGGAGACTTCTTGACCGGAGCGGACCGGCTTGAC	689
Db	457	ATGTGCTGACCTGTGTCGGAAGGTTAAAGCGCGGTTACGCGCGGCGCATGTGCTGCCG	516
Qy	690	GGCGCAGAAGGGCTCCGCGCGGACGGCTGCAACCGCTCGGATCTCTCCATCGCGATGCA	749
Db	517	GCCACCGAAGCTGCTGAGTGGGCTGAGCGGATTAACCTGGCGCGGAAGAGGG	576
Qy	750	CTGCGCTGTGTCACCGGACCTCCGCCATGACCGGGAATCGCGCTGGTGAATGCTCAGCC	809
Db	577	CTGCGCTACTCAACGCGCACCCAGGCTCCGAGCCGCTTTTGGCCCTGCGCGGTCTGTTCCAA	636
Qy	810	TGCGCCATCTCGGCAACTGGCGGCTGGCTGTAGCGGCCCTGCTTGGGAATGCTGAGA	869
Db	637	GCGGAGATCTGTTCGCTCCCGGCTCGTCTGCGGGGCGTTTAAACCGAAGCGCGCGTG	696
Qy	870	GGCGGACGAGGATGAGCGCGCGGCACTGTCCGACCTGCGCGCGCATCCCGGACAGAG	929
Db	697	GGCTCAGCTCGTTCGTTTATGCGCGCATCCATGAATGCGCGCCAGCGCGCGGAGATT	756
Qy	930	GACCCGACGAGGCTCGCGCGCCGCTGACGGCAGCGCGCGGCTGGTCCGCGACGTC	989

QY 213 GACAGGCCATGCGTGGCGAGGGCGCGCGGGGATGTGCTTGGCCCTCCGGCGCGC 272
Db |||||
QY 49 GACAGCTTATCGCGTTGCGCGCCAGAAAGCCGCATCAGCATTTCTCCGCAAGTACTT 108
Db |||||
QY 273 GACCGTGGCGTGGTCCGAGCGCGGCTCGGCGCTGTATCCCGAGGCGCGGCATGTC 332
Db |||||
QY 109 GAGAACTGGCTTCGTCGCGAGCACATATCGATGCACTAGCATCCGCTGATATCCCGGTT 168
Db |||||
QY 333 TAGCGACTGACAAACCGGCTTCGTTCCCTTGGCGAACCGCTGATCTCAGGTGAGAAATGTC 392
Db |||||
QY 169 TATGGCATTTCAACCGGCTTTGGCGGCTTGGCAACCGCCACATCGCACCCGAGGATCGC 228
Db |||||
QY 393 CGAAGCTCGAGGCGCAATCTGTGTCATCTGTGCGAGCGCGGTGGGACCGGTGCTTGAC 452
Db |||||
QY 229 GCCAAGCTGAGCGCTCCCTCATCCGTTCCACGCTGCTGCGCATGGGTGAACCGGTGAG 288
Db |||||
QY 453 TGGACGAGCGCGCGCCCATGGTTCTGGCGCGCTGTGTGTCGATCGCTCAGGAGCGCTCC 512
Db |||||
QY 289 CGCGAAGTGTGCGCGCATTTGATGTTCTTGGCTGCAAGACCCCTGGGCTTCGGCGCGCAG 348
Db |||||
QY 513 GGTGCCAGCGAGGAGCACATCGCTCGGCTGATCGACCTGCTCAATTCGAGCTCGCTCCG 572
Db |||||
QY 349 GCGTTCGCCCGGTTGCTTGTGAGACCATGTCGCAATGCTCAATGAGGCATCATCTCCG 408
Db |||||
QY 573 GCCGTTCCAGCGCGCGCACGTTGGCGGCTCGGCTGACCTGACACCGCTTGGGCGATATG 632
Db |||||
QY 409 GTAGTCGCGAATAGCGTTCACTGGGCTGCTCGGTTGACTTGGCTCGCTGCTCGACTGC 468
Db |||||
QY 633 GTGCTGTGCTTCAGGCGCGGAGACTTCTGTGAGCGGAGCGGCGCGGCTTGAAGCGC 692
Db |||||
QY 469 GCATTAGTGTGATGGCGGAGGCGAAGCACCGATGCCCCACGCGACATCCCGCGGTA 528
Db |||||
QY 693 GCAGAGGGGCTCGCGCGCGACGCGTGCAACCGCTGATCTCTCCATCGCATGCGACTG 752
Db |||||
QY 529 CCGGAAGTGTTCGCGAGGCGGATGACCCCTGTGAACTGCGCAGAAAGGAGGCGCTG 588
Db |||||
QY 753 GCCTGTGTCACCGGACCTCCGCATGACCGGATCGCGTGGTGAATGCTCACGCGCTGC 812
Db |||||
QY 589 GCTCTGTGTCACCGCACCGACGCGATGCTCGGCGAGCTGATCATGCAATTTGGCGGACCTC 648
Db |||||
QY 813 CGCCATCTCGGCAACTGGGCGGTGGGTGACGGCGCTGCTTCCGGAATGCTGAGAGGC 872
Db |||||
QY 649 GATGAGCTGCTGACATCGCCGATGCGCCGATGCGCGCCATGAGCGCTTGAAGCCCACTGGGC 708
Db |||||
QY 873 CGGACCGAGGATGSGCGCGGCACTGTCCGA---CCTGCGCGCATCCCGACAGAG 929
Db |||||
QY 709 ACCGATCAGGATATTCGCGCGAAGTGTGCAAGAACCTGCGCGCCGACCCAGGCGGCGG 768
Db |||||
QY 930 GACCGCGACGAGGCTGCGCGCGCGTGGAGCGGCGAGCGCGGCTGGTTCGGGCACTGTC 989
Db |||||
QY 769 CGCAGCGCCAGAAATGTTGCGCTTCTGCGCGACTCGCCCAATTTGCTCGC----- 823
Db |||||
QY 990 ATTGCGAGCGGAGGCTCGAAGCGCGGATATCGGAGCGGAGCGGAGCGGCGGCGAGAT 1049
Db |||||
QY 824 -----ATCGCGAGGAGAGCGCGC-AGTGCAGGAT 852
Db |||||
QY 1050 GCCTACAGCTGCGCTCGCGCTCGCAGGTTCTCGGGCGGCTTCGACAGCTCGCATGG 1109
Db |||||
QY 853 GCCTACTGCTGCTGCTGCGCGAGGTCACGCGCGCGCGCGCGGACCAATTTGCTCAT 912
Db |||||
QY 1110 CATGACCGGCTGTGACGATCGAGCTGAACCGCGGTGACCGCAATCCGGTGTTCGCGCC 1169
Db |||||
QY 913 GCCCGCTGCTGCGCCACCGCGAATGGTGTGCGGCAATTGACAACTGTGGTGTGCGCC 972
Db |||||
QY 1170 GATGGAGGCTGCCCGCTGCGACCGGGGCAATTTCAATGGGCGAGCATGTGGCGCTGAGC 1229
Db |||||
QY 973 AGCGGCGAAGTGACTTC-----CAACGGCAACTTCCACGGCGACCGGTAGCTACGTG 1026
Db |||||
QY 1230 TCCGATGCGCTGCCACGCGCGTTCACGCTTCTGGCGGCGCTTGGAGCGCGCAGATTGCA 1289
Db |||||
QY 1027 CTGGACTTCTTGCCATCGCGTGGCGGACCTCGGCTCTATCGCGGAGCGCGCCGAC 1086
Db |||||
QY 1290 CGTCTGACAGATGAAGGCTGAACCGTGGGCTGCCCCCTTCTCCACCGGGGCGCGCC 1349
Db |||||

Db 1087 CGCATGCTCGACCGACCGCTCCCGGACCTCCCGGCATTTCTGGCGGACGATCCGGGT 1146
QY 1350 GGGTTGAATTCGGCTTCATGGCGGCACAGGTGACGGCGACCGGCTCTCTGGCGGAGATG 1409
Db 1147 GTGGAATCGGGCATGATGATCGCCAGTACACCCAGGCGGGCTTGGTGGCAGAAACAAG 1206
QY 1410 CGAGCCACGGGACCTCGCTCGATCCATTCGATCTCCACGAAACCGCCCAATCAGGATGTG 1469
Db 1207 CGGCTGGCAGTTCTCGCAGCGTTGACTCATCCCATCTCTCGGCCATGAGGAGACCAC 1266
QY 1470 GTCTCGCTTTGGACCATCGCGCGCGC 1496
Db 1267 GTTTCCTGGCTGGCATCGCGCGCGC 1293

RESULT 7
US-09-833-745-12
; Sequence 12, Application US/09833745
; Patent No. 6939541
; GENERAL INFORMATION:
; APPLICANT: ROBERTS, JOSEPH
; APPLICANT: SETHURAMAN, NATARAJAN
; APPLICANT: MACALLISTER, THOMAS
; TITLE OF INVENTION: CLONING, OVEREXPRESSION AND THERAPEUTIC USE OF
; TITLE OF INVENTION: BIOACTIVE HISTIDINE AMMONIA LYASE
; FILE REFERENCE: 078728/0106
; CURRENT APPLICATION NUMBER: US/09/833,745
; PRIOR FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/197,770
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 12
; LENGTH: 1536
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic expression
; OTHER INFORMATION: vector sequence
US-09-833-745-12

Query Match 8.3%; Score 153.8; DB 3; Length 1536;
Best Local Similarity 48.9%; Pred. No. 1.4e-21;
Matches 629; Conservative 0; Mismatches 607; Indels 51; Gaps 6;

QY 213 GACAGGCCATGCGTGGCGAGCGCGCGGCTGTGCTTGGCCCTCCGGCGCGC 272
Db 49 GACGAGCTTATCGCGTTGCGCGCCAGAAAGCCGCATCAGCATTTCTCCGCAAGTACTT 108
QY 273 GACCGTGGCGTGGTCCGAGCGGCTCGGCGCTGTGTCATCCGCGAGGCGCGCATGTC 332
Db 109 GAGGAATGCGCTTCGCTCGGAGCAGATATCGATGCACTAGCATCCGCTGATATCCCGGTT 168
QY 333 TAGCGACTGACAAACCGGCTTCGTTCCCTTGGCGAACCGGCTGATCTCAGGTGAGAAATGTC 392
Db 169 TATGGCATTTCAACCGGCTTTGGCGGCTTGGCAACCGCCACATCGCACCCGAGGATCGC 228
QY 393 CGAAGCTCGAGGCGCAATCTGTGTCATCATCTGGCGAGCGGCTGGGACCGGTGCTTGAC 452
Db 229 GCCAAGCTGACGCGCTCCCTCATCCGTTCCACGCTGCTGCGCATGGGTGAACCGGTGAG 288
QY 453 TGGACGAGCGCGCGCCCATGGTTCTGGCGCGCTGTGTGTCGATCGCTCAGGAGCGCTCC 512
Db 289 CGCGAAGTGTTCGCGCATTTGATGTTCTTGGCTGCAAGACCCCTGGCTTCCGG---CCGC 345
QY 513 GGTGCCAGCGAGGAGCACATCGCTCGCCCTGATCGACTGCTCAATTCGAGCTCGCTCCG 572
Db 346 AGCTTCCCGCGGTTGCTTCTGAGACCATGGTGGCATGCTCAATCGCAGGCATCACTCCG 405
QY 573 GCCGTTCCAGCGCGGACGCGTGGCGGCTCGGCTGACCTGACACCGCTTGGCATATG 632
Db 406 GTAGTCCGGAATAGGTTCACTGGGCTGCTCCGCTGACTTGGCTCCGCTCCGCTCGCACTGC 465

```
QY 633 GTGCTCTGCTCCAGGCGCGGAGACTTCTTGGACCGGAGCGGCGGCTTGACGGC 692
Db 466 GCATTAGTGTGATGGGCGAGGGGCAAGCACCAGATGCCAGCGGACATCGGCCGGTA 525
QY 693 GCAGAGAGGCTCCGGCGCGGAGCGGTGCAACCGCTCGATCTCTCCCATCGGATGCACTG 752
Db 526 CGGAACTGTTGCGGAGCGGAGTGGACCTGTGCACTGGGAGGAAAGGAGGCGCTG 585
QY 753 GGGCTGTGTAACGGGACCTTCCGCCATGACCGGATGCGCGTGGTGAATGCTCAOCCCTGC 812
Db 586 GCTCTGTGTAACGGGACCGGACGATGCTCGGCCAGCTGATCATGGCATTTGGCGACCTC 645
QY 813 CGCCATCTCGCAACTGGGCGGTGGGTGACGGGCGCTTGGCGATGCTGAGAGGC 872
Db 646 GATGAGCTGTGGACATCGCGATGCCACCGCGCCATGAGCGTTGAAGCCAGCTGGGC 705
QY 873 CGGACCGAGGATGGCGCGGCACTGTCCGA---CCTGGCGCGCATCTCCGGACAGAAG 929
Db 706 ACCGATCAGTATTCCGGCGCAGACTGCAGNACCACTGGCGCGCACCGAGGCCAGGGC 765
QY 930 GACGCCGAGGAGCTCGCGCGCGGTGGAGCGGAGCGCGGGTGGTCCGGCACGTC 989
Db 766 CGCAGCGCCAGAACATGTTGCGCTTCTGCGCGACTCGCCAAATTTGTTGCTCGC----- 820
QY 990 ATTGCCGAGCGAGGCTCGACCGCGGCGATATCGGGACGGAGCGGAGCGGGCAGGAT 1049
Db 821 -----ATCGCGAGGGAGACGGCGG-AGTGAGGAT 849
QY 1050 GCCTACAGCTCGCTCGGCTCCGAGGTTCTCGGGCGGGCTTCCGACACGCTCGCATGG 1109
Db 850 GCCTACTCGCTCGGTTGCTCGCGGAGGTACCGGCGCGCGCGCGGACACCATTTGCTCAT 909
QY 1110 CATGACCGGGTGTGACGATCGAGCTGAACCGGCTGACCGCAATTCGGTGTTCGCGCC 1169
Db 910 GCCCGCTGGTCCGACCGCGGAACTGGCTCGCGGCAATTGACAACTGTTGGTGTGCGCC 969
QY 1170 GATGGCAGCTGCGCGCGCTCGACGGGGGCAATTCATGGGCGAGCATGTGGCGTGAAG 1229
Db 970 AGCGCGGAAGTGACTTC-----CAACGGCAACTTCCACGGCGCACCGGTAGCCTACGTG 1023
QY 1230 TCCGATGGCTCGCACCGCGCTCACCGTTCTGGGGCGCTTGGGAGCGCCAGATTGCA 1289
Db 1024 CTGACTTCTTGTGCATCGCGTGGCGGACCTCGGCTCTATCGCGAGCGCGCACCGAC 1083
QY 1290 CGTCTGACAGATGAAGGCTGAACCGTGGGCTGCGCCCTTCTCTCACCGGGCGCCCGCC 1349
Db 1084 CGCATGCTCGACCCAGCGCGCTCCGCGACCTGCGCGCATTCCTGGCGGACGATCCGGGT 1143
QY 1350 GGGTTGAATTCGGCTTCATGGGCGGACAGGTGACGGCGGACCGCGCTCTGGCGGAGATG 1409
Db 1144 G---TGGACTCGGGCATGATGATCGCCAGTACACTCAGGCGCGCTTGGTGGCAGAAAC 1200
QY 1410 CGAGCCAGGACCTGCTCGATCCATTCGATCTCCAGNACGCGGCCAATCAGGATGTG 1469
Db 1201 AAGCGGCTGGCAGTTCTTGGAGTTGACTTCATCCCATCTCGGCCATGACAGAAAGACAC 1260
QY 1470 GTCTCGCTTGGGACCATCGCGCGCGC 1496
Db 1261 GTTTCCTGGGCTGGCATGCGCGCGC 1287
```

RESULT 8

```
US-09-252-991A-16124/c
; Sequence 16124, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
```

```
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16124
; LENGTH: 1557
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16124
```

Query Match

Best Local Similarity 6.8%; Score 125.4; DB 3; Length 1557;

Matches 714; Conservative 0; Mismatches 706; Indels 75; Gaps 9;

```
QY 225 GCCGTGGCGAGCGCGCGCGGATTTCTTCCCTTCGCGCGCGGACCGGTCGCGT 284
Db 1506 GCGGTGGCGCGGACCGGTGGCGCGCTGGAATCTGTGGCGCGCGCTGGGACGATCGAC 1447
QY 285 GGTTCGAGAGCGCGCTCGGCGCTGTCTATCCGAGAGCGCGCCATGTCTACGAGCTGACA 344
Db 1446 AACGCCCGCGCATCTGTCCGTATCTGTCCCAACGCGGAGCGTGTCTACGGCATCAGT 1387
QY 345 ACCGCTTCGCTCCCTTGGCAACCGCTGATCTCAGTGAAGATGTCGGAACGCTGCGAG 404
Db 1386 ACCGCTTCGCGCGCTGTGGAGCTCTGTCTGGAAGCGGAGCAGCTCGCCGAGCTGCG 1327
QY 405 GCCAATCTTGTTCATCATCTGGCCAGCGCGTGGGACCGGTCGTTGATGACGACGCGCG 464
Db 1326 CGCAACACCTTGTCTCAGCCATGCTCGGGGTGGCGGAGCGCTGCGGACGACGAGACC 1267
QY 465 CGCGCATGTTCTGGCGCTGTGTGTGATTCGCTAGGAGCGCTCGGTCGCGAGCGAG 524
Db 1266 CGGCGATCATCTGTGCGCGCTGCGCAACTACAGCCAGGCGCAATGTCGGGCTCGACCGT 1207
QY 525 GGGACCATCTGCTCCCTGATCGACCTGCTCAATTCGAGCTCGCTCGGCGCGTTCCCGACG 584
Db 1206 TCGTGTGGAGAGCTGTGGCGCTGCTCAACACAGGCAATACCCCGAGGTGCGCGCC 1147
QY 585 CGCGCACGTTGGCGCGCTCGGCTGACCTGACACCGCTTGGCGCATATGTTGCTCTGCGCTC 644
Db 1146 CAGGCTCGGTGGCG-----TACCTGACCCACATGGCGGACGCTCGGCGATCGCCCTG 1096
QY 645 CAGGCGCGGAGACTTCTCGGACCGGAGCGGCTTGAAGCGGCTTGAAGCGGAGAGGGCTC 704
Db 1095 CTCGCGCATCGCGGAGGTGAGTACCGGCGGAGCGCTGTCGCGCGCGCGCGC---CGCGTTG 1039
QY 705 CGGCGGAGCGCTGCAACCGCTCGATCTCTCCCATCGCGATGCACTGGCGCTGTTCAAC 764
Db 1038 GCGCGGAAGGCTTGGCGAGCGTGGCTGGGAGCCAAAGGACGCGGCTCTGCTTGTGTTCAAC 979
QY 765 GGGACCTCCGCGCATGACCGGAGTCCGCTGTGTAATGCTCACGCTTGGCGCATCTCGGC 824
Db 978 GGCAGCGGTGTCATGACCGGCTCGCTGCTGGCCCTGGAGCGCGGAGCGCTGCGG 919
QY 825 AACTGGCGGTGGGTTGACGGCGCTTCTGCGGAATGTCTGAGAGCGCGGACCGAGGCA 884
Db 918 CAGTGGGCGGACGTTGATCGGGCGGATGAGCTTCGAGGCGCTTGGCGCGCAACTGGCGCGC 859
QY 885 TGGGCGGCGGACTGTCCGACCTCGGCGCGCTCCGAGCAGAGGAGCGCGGAGCGGAGG 944
Db 858 TTGACCGCGGAGATCTGTTGGCTCAAGCCGATCTCCCGGATGACGCGGCTCGCCGTCAAT 799
QY 945 CTGCGCGCGCGTGGAGCGGCGCGGCTCGGCGACGCTATTTCGCGAGCGGAGG 1004
Db 798 CTGCGGCTTGTGCTGGCTGGCAGCCAGTACTGAGA----- 762
QY 1005 CTCGACCGCGCGATATCGGAGCGGAGCGGCGGCGGAGGATGCTTACAGCTGCGC 1064
Db 761 ----ACGCGCGGCGATCCGCGC-----CCAGGATGCGCTGAGCATCCGC 721
QY 1065 TGGCTCGCGAGGTTCTCGGGCGGCTTCGACACGCTTCGATGCGCATGCGGCTGCTG 1124
```


Db 720 TCGATACCGCAGATCCATGCGCGCTGCCGTGATCCAGTTGGCCCAACGCGGCGCGCAGATA 661
QY 1125 ACGATCAGCTGAACGCGGTGACCGACAATCCCGTGTTCGCG---CCGATGGCAGGTG 1181
Db 660 GAGACCGAATCAACTCGGCACCGACAACCCGCTGCTGTGGGTACCGCGAAGCCTAC 601
QY 1182 CCGCCCTGCACGGGGCAATTTTCATGGGCGCAGCATGTGGCGCTGAGCGTCCGATCGCGCTC 1241
Db 600 CGGTGTGATCCACGGGCAATCCCAACGCGGAATCGGTGGCGATGGCCGCGACCTGCTG 541
QY 1242 GCCAGGCGCTCACCGTTCTGGCGGGCTTGGCGAGCGCCAGATTGACGCTGTGACAGAT 1301
Db 540 GCGATCCCGTGGCGAGCTGGCGGGGTGCGCGAGCGGGCCCTGGATCGCCTG---GTC 484
QY 1302 GAAAGGCTGAACCGTGGGCTGCCCCCTTCTCTCCACCGGGGCGCCCGCGGTTGAATTC 1361
Db 483 AACCGCTGTGACGGCCCTGCGGCGTTCCT---GGTGGCAAGCCGGAGTCAACTCG 427
QY 1362 GGCTTCATGGGCGCACAGGTGAACGGCGACCGCGCTCCTGGCCGAGATGCG---AGCCACG 1418
Db 426 GGGATGATGATACCCAGTAGCTGCGCGCTCCCTGGCTGGCGAGAACCGCCAGCTGGCG 367
QY 1419 GGACCTGCTCGATCCATTGATTCACGAATCCAGAACCGCCCAATCAGGATGTGCTTGCCTT 1478
Db 366 CAGCCGCGGTGGTGCACAACCTTGCTCACCTCGCGCTCCAGGAGGACCACTGAGCCTC 307
QY 1479 GGAACCATCGCGCGCTCTGCGCGAGAGATCGACCGCTTGGGCGGAGATCCTTGG 1538
Db 306 GGCACAGTGGCGGCTCAAGCTTGGCAGGCGCTTGAGAACCTCCGCGGATCCTCGCC 247
QY 1539 ATCTCTGCTCTGTCTTTCACAACTGCGGAGTGGCGCTGCGCGACGCGCTTAGACGGG 1598
Db 246 ATGAGTACCTGTGGCGGCCAGGCTTCGAATTCCTGCCCGCAACGCTTGGGCCAG 187
QY 1599 GTGTCTCCCGGGGAAAGCTGTGTGACGGCCCTGCGCAGCAGTTCGCCGCGCTTGAG 1658
Db 186 GGCACCGCGCGCTGGGCAT-----CTGCGCGAGCGCGTGCCTCGCGGAC 136
QY 1659 ACGACCGGCGCTGGGACAGGAATTCGCGCTTGTCTACGCACTCTTGGACG 1713
Db 135 ACCGACCGTGGTGTCCCGAATCGCAGCGCGCGCCCATCTCTCGCGCAGC 81

RESULT 9
US-09-252-991A-16478
; Sequence 16478, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16478
; LENGTH: 1563
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16478

Query Match 6.8%; Score 125.4; DB 3; Length 1563;
Best Local Similarity 47.8%; Pred. No. 5.9e-16;
Matches 714; Conservative 0; Mismatches 706; Indels 75; Gaps 9;

QY 225 GCGTGGCAGCGCGCGCGGATGTGCTTGGCCCTCCGCGGCGGACCGGTGCGGT 284
Db 91 GCGGTGGCCCGCAGCAGGTGCGCGCTGGAATCTGCGCGCGCGCTTGGCAGCGATCGAC 150

QY 285 GCCTCCGAAGCGCGCTCGCGCTGTATCCGCGAGCGCGCCATGTCTACGACTGACA 344
Db 151 AACGCCCGGCCATCGTTCGCCGTATCGTCGCCAAACGCGAGCGTGCCTACGCATCAGT 210
QY 345 ACCGGCTTCGGTCCCTTTCGAAACCGCCTGATCTCAGGTGAGAAATGTCCGAAACGCTGCAG 404
Db 211 ACCGGCTTCGGCGCGCTGTGCGACGTCCTGCTGGAAGCGGAGCAGCTCGCCGAGCTGTG 270
QY 405 GCGAATCTTGTCCATCATCTGGCGCAGGGGCTGGGACCGGTGCTTCACTGGACGAGGG 464
Db 271 CGCAAACCTTGTCTAGCCATGCTCGGGGTGCGGAGCCGCTGCGCGACGAAACAGACC 330
QY 465 CGCGCATGTTCTGCGCGGTCTGTGTGATCGCTCAGGGAGCCTCCGCTGCGCAGCGAG 524
Db 331 CGGGCGATCATCTGTGCGCGCTGCGCAACTACAGCACAGGGCAAGTCCGGGCTCGACCGT 390
QY 525 GGAACCATCGCTCGCTGATTCGACCTGTCTCAATTCGAGCTCGCTCCGGCCGTTCCAGC 584
Db 391 TCGCTGTTGAAAGGACTGTGTGCGCTGCTCAACACCGCATTAACCCCGCAGGTGCGCGGC 450
QY 585 CGCGCACCGTGGGCGCTCGGTGACCTGACACCGCTTGGCGATATGTTGCTCTGCCTC 644
Db 451 CAGGGCTCGGTGGC-----TACCTGACCCACATGGCGCACGTGCGCATTCGCCCTG 501
QY 645 CAGGGCGGGGAGACTTCTTGACACCGGAGCGGCTTGACGGCGCAGAAAGGGCTC 704
Db 502 CTCGGCATCGCGAGGTGAGCTTACCGCGCAGCGTCTGCGCGCCCGC---CGCGTTG 558
QY 705 CGCGCGGACCGGTGCAACCGCTCGATCTCTCCATTCGCGATGCACTGGCGCTGTGTCAAC 764
Db 559 GCGCGGAAAGGCTGCGGACGCGTGGCGCTGGGAGCCAAAGGACGGGCTCTGCTGTGTCAAC 618
QY 765 GGGACCTCCGCCATGACCGGATCGCGTGGTGTGATGCTCAGCCTGCGCCCATCTCGGC 824
Db 619 GGCACCGCTGATGACCGGCTCGCTCGCTGCGCCTGCGACGCGCAGCGCCTGGCG 678
QY 825 AACTGGCGGTGCGGTGACGCGCTCTTGGCGAATGCTGAGAGCGCGACCCGAGGCA 884
Db 679 CAGTGGCGCGAGTGTGATCGGGGCGATGAGCTTCGAGGCCCTTGGCGGCCCAACTGGCGGCG 738
QY 885 TGGGCGCGCGCACTGTTCGACCTGCGGCGCGCATTCGCGGACAGAAAGACCGCGCAGCGAGG 944
Db 739 TTAGACCGCGAGATCTTGGCTCAAGCGCGATCCCGGATGCGAGCGGTGCGCGCTCAAT 798
QY 945 CTGCGCGCGCGGTGACGCGCAGCGCGGTGGTCCGCGACGTCATTCGCGAGCGGAGG 1004
Db 799 CTGCGAGCCTTGTGCTGGCAGCCAGGTACTGGAGA----- 835
QY 1005 CTCGACGCGCGGATATCCGGACCGAGCGCGAGCGGGGCGAGGATCGCTACAGCCTGCGC 1064
Db 836 ----ACGCCCGGGCATCCGAC-----CCAGGATGCCCTGAGCATCCGC 876
QY 1065 TGGCTTCGCGAGTTCCTCGGGGCGGGCTTCGACACGCTCGCATGGCATACCGGGTGTGCTG 1124
Db 877 TCGATACCGCGCAGATCATTCGCGCTGCTGCGTGCAGTTCGGCCACCGCGCGCGCGCAGATA 936
QY 1125 ACGATCGAGCTGAACGCGGTGACCGGACATCCGTTGTTTCGCG---CCGATGSCAGGTG 1181
Db 937 GAGACCGAACTCAACTCGGCCACCGACAACCCGCTGCTGTGGGTACGCGCGGAAGCCTAC 996
QY 1182 CCGCGCCTGCACGGGGGCAATTTTCATGGGCCAGCATGTGGCGCTGACGCTCCGATGCGCTC 1241
Db 997 CGGTGTGATCCAGGCCAATCCCGACGCGGAATCGGTGCGCATGGCGCGCGACCTGTGCTG 1056
QY 1242 GCCAGGCGCTCACCGTTCTGGCGGCGCTTGGCGAGCGCCAGATTGCACTGTCTGACAGAT 1301
Db 1057 GCGATCGCGCTGCCGAGCTGGCGGGGTGCGCGAGCGCGCTGATCGCCTG---GTC 1113
QY 1302 GAAAGCTGAACCGTGGGCTGCCCTTCTCCACCGGGGCGCCCGCGGTGTAATTC 1361
Db 1114 AACCCGCTGTCAGCGGCTGCGCGGTTCCT---GGTGGGCAAGCCGGGAGTCAACTCG 1170
QY 1362 GGCTTCATGGGCGCACAGGTGACGGCGACCCGCGCTCTCTGGCGGAGATGCG---AGCCAGC 1418

Db 1171 GGGATGATACCCAGTACGTGCGCGCTTCTGCTGCGGAGAACCGCCAGCTGGCG 1230
Qy 1419 GGACTGCTCGATTCATTCGATCTCCACGAAACCGCCGCAATCAGGATGTGTTCTCGCTT 1478
Db 1231 CAGCGCGGGTGGTTCGAACTTCGTACCTCGCGCTCCAGGAGGACCACTTACGCTC 1290
Qy 1479 GGGACCATCGCGCGCTCTCGCGGAGAGATCGACGTTGGGCGGAGATCTTGGG 1538
Db 1291 GGCACCAAGTCCGCGCTCAAGCTTTGGCAGGCGCTTGGAGAACCTCCCGCGGATCTCGCC 1350
Qy 1539 ATCTCTGCTCTCTCTTTGCAACAGCTGCGAGCTGCGCTGCGCAGCGGCTAGACGGG 1598
Db 1351 ATCAGTACCTGCTGGCGCGCCAGGCTTCGAATTCCTGCGCCGCAACGCTTCGGCAG 1410
Qy 1599 GTGTCTCCCGCGGAGAAAGCTGGTGAAGCCCTGCGCGAGCAGTTCCCGCGCTTTGAG 1658
Db 1411 GGCACCGCGCGCTGCGGCTT-----CTGCGCGAGCGCTGCGCGGTACGAC 1461
Qy 1659 ACGGACCGCGCTGGGACAGGAAATTTGCGCGCTTGTACGCACTCTTGCAGC 1713
Db 1462 ACCGACCGTGGCTGGCTCCGACATCGCCAGCGCGCGGCACTCTCTCGCGAGC 1516

RESULT 10

US-09-624-693A-16
; Sequence 16, Application US/09624693A
; Patent No. 6355468
; GENERAL INFORMATION:
; APPLICANT: Yoshida, Roberta
; APPLICANT: Kootstra, Anna
; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
; TITLE OF INVENTION: Using Same
; FILE REFERENCE: 29479/500NSC
; CURRENT APPLICATION NUMBER: US/09/624,693A
; CURRENT FILING DATE: 2000-07-24
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 2787
; TYPE: DNA
; ORGANISM: Rhodotorula mucilaginosa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (646)..(2784)
US-09-624-693A-16

Query Match 6.0%; Score 110.4; DB 3; Length 2787;
Best Local Similarity 46.6%; Pred. No. 5.9e-13;
Matches 514; Conservative 0; Mismatches 566; Indels 24; Gaps 4;
Qy 483 CGTCTGCTGATCGCTCAGGGAGCTCGGTCGCCAGCGAGGGACCATCGCTCGCTG 542
Db 1177 CGTGTCAACTCGCTCATCTCGGGTCACTCGGGTTCGCGATGCTGCTTCGAGGCCCTC 1236
Qy 543 ATCGACTGCTCAATTCGAGCTGCTTCCGCGCGTTCCTCCAGCGCGACAGTGGGCGG 602
Db 1237 ACCAACTTCTCAACACGCGCATCACCCGATCGTCCCGCTTCGAGGACCACTTCGGG 1296
Qy 603 TCGGTAACCTGACACCGCTTGGCGATATGTTGCTCTGCTTCCAGGCGCGGAGACTTC 662
Db 1297 TCGGGGACCTTTTCCCGCTCTTACATCGCGCTTCGATCACTCCGCGCACCGGACTCG 1356
Qy 663 CTGG--ACCGGACGCGGAGCGGCTTACGCGGCGAGAGGCTCCGCGCGGACGCGCTG 719
Db 1357 AAGTCCACGTCGACGCGAAGATCATGTCCGCCAGGAGGAGTTCGGCTCAAGGGTCTT 1416
Qy 720 CAACCGCTCGATCTCTCCCATCGGATGCACTGGCGCTGGTTCACAGGAGACTTCGCCCATG 779
Db 1417 CAGCCCGTCTGCTCGTCCGAGAGGAGTCTCGGTCTCTGCTCAACGCGCAGCGCGTCTCC 1476
Qy 780 ACCGGGATCGGCTGGTGAATGCTCAACGCTTCGCCCATCTTCGGCAACTCGGCGGTGGG 839

Db 1477 GCCTCGATGCGGACGCTGGCCCTCACCGACGCAACAGTCTCTCTGCTCTCTCGCACAGCGC 1536
Qy 840 TTGACGCGCTTGTGTTGGGAATGTCTGAGAGGCGGACCGAGGCGATGGCGCGGCACTG 899
Db 1537 CTCACTGCTTACTGTGCGAGGCAATGGTCGGACACGCGCGCTGTTCCACCACTTCTC 1596
Qy 900 TCGACACT---GGCGCGCATTCGCGGACAGAGACGCGCGAGGAGGCTGCGGCGCGCG 956
Db 1597 CAGCAGCTCACGCGCGCTTACCGCGACCCAGATCGAGGTGGCGGCAACATCCGGACTCTT 1656
Qy 957 GTGAGCGGACGCGCGGCTGTCGCGACGTCATTCGCCGAGCGAGGCTCGACGCGCGC 1016
Db 1657 CTGAGGCGAGCAAGTACGCCGCTCCACCGAGACTGAAGTCAAGTCAAGGTCAGAGGACGAG 1716
Qy 1017 GATATCGGACGAGCGCGGCGGAGGAGTGCCTTACAGCCTTGCCTGCGCTCCGCGCAG 1076
Db 1717 GGCATCTCTCA-----GGCAGGACCGGTACCGCTCGCTGCTCGCGCGCAG 1761
Qy 1077 GTTCTCGGCGCGGCTTTCGACACGCTCGCATGGATGACCGGGTGTGAGATCGAGCTG 1136
Db 1762 TGGCTCGGTCCTTGTGTCAGCGACATGATTCACGCTCACGCTGTCTCTCTGCTCGAGGCT 1821
Qy 1137 AACCGGTGACGACCAATCGGTGTTTCCGCCGATGGGAGCGTGGCCGCTGCACGCGG 1196
Db 1822 GGTGAGTCGACCAACCGCAACCGCTGATCGAATTCGAGAACAGATGACCCACCATGGC 1881
Qy 1197 GGCATATTTTCATGGCGCGCATGTGGCGCTGACGCTCGATCGCTGCGCACGCGCGCTCACC 1256
Db 1882 GGAGCTTTCATGGCGAGCAGCGTTCGGAACACCATGGAGAGACTCGCTCGCGCTCGCG 1941
Qy 1257 GTTCTGGCGGCGCTTGGGAGCGCGAGTTGACGCTGTGACAGATGAAGGCTGAACCGT 1316
Db 1942 CTGATGGCAAGGTCACTTACTCAGCTCACGAGATGCTCAACGCGCGCATGAACCGG 2001
Qy 1317 GGGCTGCGCGCTTCTTCCACGCGGCGCGCGGGTGAATTCGCGCTTTCATGGGCGCA 1376
Db 2002 GCGCTTTCGCTGCTGCTGCGGAGGACCTTCCCTCTTATCACTCAAGGGGTCTC 2061
Qy 1377 CAGGTGACGCGGACCGCTTCTGCGCGAGATCGGAGCC---ACGGGACCTGCTCGATC 1433
Db 2062 GACATTCGCTGCGCGCGCTTACACTTCGAGCTCGTTCACCTTCGCAACCGGTTTCGACC 2121
Qy 1434 CATTCGATCTCCAGAACCGCGCAATCAAGATGTGTCTGCTTGGGACATCGCGCGG 1493
Db 2122 CAGTCCAGCGCGCGGAGATGGGCAACCAAGGCGCATCAACTCGCTCGCGCTCATCTCGGCC 2181
Qy 1494 CGCTCTGCGCGGAGAGATCGACCGTTCGGGCGGAGATCCTTGGATCCTCGCTCTCTGT 1553
Db 2182 CGCGCACCGCGCGGAGGAGACGAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2241
Qy 1554 CTTGCAACAGCTCGGAGGCTCGCGC 1577
Db 2242 GTCTCCAGGCGCTCGACCTCGCGC 2265

RESULT 11

US-09-252-991A-16368
; Sequence 16368, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16368

QY 1254 ACCGTTCTGGCGGCTTTCGGAGCGCCAGATTGACGCTTGACAGATGAAAGGCTGAAC 1313
DB 1312 GCCCTGATCGGCAAGCTCAATTACGAGCTACACGAGATGCTCAACGCCGCGATGAAC 1371
QY 1314 CGTGGGCTGCCCTTCTCTCCACCGGGGCCCGCGGGTTGAATTCGGGCTTCATGGGC 1373
DB 1372 CGCGCTTNCNTCTGCTCGTCCGAGGACCCNTCGCTCTCTCTATCACTGCAAGGCG 1431
QY 1374 GCACAGGTGACGGGACCGCGCTCTGCGCGAGATGCGAGCC---ACGGAGCTTGCCTCG 1430
DB 1432 CTCGACATTGCGCGNGCNGCNTACATTTCGGAGCTCGGNCACCTTTCGCAACCCGCTNACG 1491
QY 1431 ATCCATTTCGATCTCCACGAACCGCCCAATCAGGATGTGCTGCTTTGGGACCATGCCC 1490
DB 1492 ACCACGTCACCGCGNGAGATGGGCAACAGGCCGCTCAACTCGCTCGCMTACTCTCG 1551
QY 1491 GCGGCTCTGCGCGGAGAGATCGACGTTGGCGGAGATCTTTGCGATCTCTGCTCTC 1550
DB 1552 GCGCGCGCACNGCGGAGGCAACGAGCTCTTTCTCTCTCTCTCTCTCTCTCTCTCTCT 1611
QY 1551 TGTCTTGACAGCTGCGGAGCTGCGC 1577
DB 1612 TCGCTNCTCCAGGCGCTCGACCTCGC 1638

RESULT 13

US-09-765-873A-31
; Sequence 31, Application US/09765873A
; Patent No. 6521748
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BC1009 US CIP
; CURRENT APPLICATION NUMBER: US/09/765,873A
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 31
; LENGTH: 2061
; TYPE: DNA
; ORGANISM: mutant from Rhodotorula glutinis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2061)
US-09-765-873A-31

Query Match 5.6%; Score 103.2; DB 3; Length 2061;
Best Local Similarity 46.5%; Pred. No. 1.5e-11;
Matches 511; Conservative 0; Mismatches 553; Indels 36; Gaps 4;

QY 499 CTCAGGAGCTCCGGTCCGACGAGGGGACCATCGCTCGCTCATCGACCTGCTCAATT 558
DB 440 CCCCGGCCACTCGGCTGTCGCTCGCTGCTCGAGGCGCTCACCAACTTCTCAACC 499
QY 559 CCGAGCTCGCTCCGCGGCTTCCAGCGCGCGCGCTGGGCGCTGGGTGACCTGACAC 618
DB 500 ACGGCATCACCCCATCGTCCCTCCGCGCACCATCTCTGCTCGGGGACCTGTCTC 559
QY 619 CGCTTGGCGATPAGGTGCTCTGCTTCCAGGGCCGGGAGACT-----TCC 663
DB 560 CTCTCTCTATATTGACAGCGCCATCAGCGGTCAACCGACAGCAAGGTGCACTGCTCC 619
QY 564 TGGACCGGACGAGCGGCTTCAACCGCGCAGAGGCTCCGCGGAGCGGTGCAAC 723
DB 620 ACGAGGCAAGGAGAAGATCTGTACGCCCGCGAGCGGATGGCGCTTCTCAACCTCGAGC 679
QY 724 CGCTCGATCTCTCCCATCGCGATCACTTGGCGGTGCTCAACGGGACCTCCGCGCATGACCG 783
DB 680 CCGTCTGCTCTCGGCGCGAGAGGTCTCGTCTCTGTAACGGCAGCGCGCTCTCAGCAT 739

QY 784 GGATCGCGCTGGTGAATGCTCACGCTGCGCCCATCTTCGGCAACTGGGCGGTGGGTTGA 843
DB 740 CGATGGCCACCCCTCGCTCTGCACGACGCTCATCTCTCTCTCTCTCTCTCTCTCTCT 799
QY 844 CGGCGCTTGTTCGGAATGTCTGAGAGCGCGGACCGGAGCATGGCGCGGCACTGTCCG 903
DB 800 CGGCGCATGACGCTCGAAGCATGTTCGGCCACGCGCGCTCTGTTCCACCCCTCTCTCAG 859
QY 904 ACCT---GCGCGCCATCTCCGGAAGGACGCGCAGCGAGGCTGCGCGCCCGCGTGG 960
DB 860 ACGTCAACGCGCTTTCGCTTCCACCATGAGGAGGAGGTCAAGGTCAAGGACACGAGGCA 919
QY 961 ACGGACGCGCGCGGTGGTTCGGGACGCTCATTCGCGAGCGAGGCTCGACGCGCGGATA 1020
DB 920 AGGAAAGCCGCTTTCGCTTCCACCATGAGGAGGAGGTCAAGGTCAAGGACACGAGGCA 979
1021 TCGGAGCGGAGCGGAGGCGGGGAGGATGCTACAGCTCGCTGCGCTCGGCTCGGAGGTT 1080
DB 980 TT-----CTCGCGCAGGACCGCTACCCCTTGGCAGCGTCTCTCTCTCTCTCTCT 1024
QY 1081 TCGGCGCGGCTTCGACACGCTCGCATGCGATGACCGGGTCTGACGATCGAGCTGAACG 1140
DB 1025 TCGGCGCGCTCGTCAAGGCTTTCGCGCCACCGCGCTCTCTCACCATCGAGGCGCGCC 1084
QY 1141 CGGTGACGACATTCGCGGTTCGCGCGCGGATGCGAGGCTGCGCGCTCGACGCGGCA 1200
DB 1085 AGTCGACGACCGACCAACCTCTCATCGACCTCGAGAACCAAGACTTCGACACGCGCGCA 1144
QY 1201 ATTTTCATGGGCGCAGCATGTGCGCTGACGCTCCGATGCGCTCGCCACGCGCTCACCGTTC 1260
DB 1145 ATTTTCAGGCTCGCGCTGTGCGCCACACCATGAGGAGAGACTCGGCTCGGGCTCGCCGAGA 1204
QY 1261 TGGCGGCGCTTTCGAGGCGCGCAGATTGACAGCTCTGACAGATGAAAGGCTGAACCGTGGCG 1320
DB 1205 TCGGCAAGCTCAACTTCACGAGCTCACCGAGATGCTCAACCGCGCATGAACCGCGGCC 1264
QY 1321 TGCCCCCTTCTCCACCGGGGCGCGCGGGTTGAATTCGGGCTTATGGGCGGACAGG 1380
DB 1265 TCCCTCTCTCTCGCGCGGCGAAGACCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1324
QY 1381 TGACGCGCGCGCGCTCTCTGCGGAGATGCGA---GCCAGGGACCTGCTCGATCCATT 1437
DB 1325 TCGCGCTGCGCGGTACACCTCGGAGTTGGACACCTTCGCAACCTCTGACAGCGCATG 1384
QY 1438 CGATCTCCAGAACCGCGCAATCAGGATGTGCTCTGCTTGGGACCATCGCGCGCGCC 1497
DB 1385 TCCAGCGGCTGAGATGCGGACCGGCGGTCAACTGCTTGGCTCATCTCGGCTCGTC 1444
QY 1498 TCTGCGCGGAGAGATCGACCGTTGGGCGGAGATCTTGGGATCTCTGCTCTCTCTCT 1557
DB 1445 GCACGACCGAGTCCACGAGCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1504
QY 1558 CACAAGCTGCGGAGCTGCGC 1577
DB 1505 TCCAAGCATCGACTTGGC 1524

RESULT 14

US-09-627-216A-7
; Sequence 7, Application US/09627216A
; Patent No. 6368837
; GENERAL INFORMATION:
; APPLICANT: Sariaelani, Sima F
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Qi, Wei Wei
; APPLICANT: Vannelli, Todd
; APPLICANT: Gatenby, Anthony
; TITLE OF INVENTION: Bioproduction of para-Hydroxycinnamic Acid
; FILE REFERENCE: BC1009 US NA
; CURRENT APPLICATION NUMBER: US/09/627,216A
; CURRENT FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: 60/147,719

; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 7
; LENGTH: 2151
; TYPE: DNA
; ORGANISM: Rhodotorula glutinis
US-09-627-216A-7

Query Match 5.6%; Score 103.2; DB 3; Length 2151;
Best Local Similarity 46.5%; Pred. No. 1.5e-11;
Matches 511; Conservative 0; Mismatches 553; Indels 36; Gaps 4;

QY 499 CTGAGGAGCTCCGGTGCAGGAGGGACCATCGCTCGCTGATCGACCTGCTCAATT 558
DB 530 CCGCGGCACTGGCTGTCCGCTCGCTCGCTCGAGGGCTCACCACATTTCTCAACC 589
QY 559 CCGAGCTCGTCCGGGCGGCTTCCAGCGCGGCAAGGCTCCGGCGGCTCGGGTGACCTGACAC 618
DB 590 ACGGCATCACCCCATCGTCCCTCCGGCGCACCATCTCTGCTCGGGGACCTGTCTC 649
QY 619 CGCTTGGCATATGCTCTGCTCTCAGGCGGGGAGACT-----TCC 663
DB 650 CTCTCTCTACATTCGAGCGCCATCAGCGGTCAACGAGCAAGGTGCACGTCTGCTC 709
QY 664 TGGACCGGGACGGACCGCGCTTTCAGCGCGCAGAGGGCTCCGGCGGACGCTGCAAC 723
DB 710 ACGAGGCAAGGAGAGATCTGTACGCCCGGAGGAGTGGCTCTTCAACTCGAGC 769
QY 724 CGCTCGATCTCTCCCATTCGCGATGCACTGCGCTGTGTCACAGGAGCTCCGCGCATACCG 783
DB 770 CCGTCTGCTCGGCGGCAAGGAGTCTCGCTCTCGTCAACGGCACCGCGCTCTCAGCAT 829
QY 784 GGATCGCGCTGGTGAATGCTCAGCGCTGCGCGCATCTCGGCAACTGGGCGGTGGCTTGA 843
DB 830 CGATGGCCACCTCGCTCTCAGCGAGCTCACAATGCTCTCGCTCTCTCGAGTCTGCTCA 889
QY 844 CGGCTCTGCTCGGAGTGTCTGAGAGCGGACCGGACCGAGGATGGCGCGGCTGTTCCACCCCTTCTTCAACG 949
DB 890 CGGCAATGAGCTGCGGAGCGGATGGTTCGCGTCAACCGGCTCTCGCTCTCTCGAGTCTGCTCAACG 960
QY 904 ACCT---GCGGCGGCTTCCGGACAGAGGACCGCGAGGCTGCGCGCGCGGTGG 960
DB 950 ACGTCAAGCGCTTCAACCGGACGAGTCAAGTTCGCGGAAACATCCGCAAGCTCTCTG 1009
QY 961 ACGGACGCGCGGCTGCTCGGACAGTCAATTCGCGAGCGGAGCTCGACCGCGGAGTA 1020
DB 1010 AGGGAAGCGCTTTGCTGTCTCCATGAGGAGGAGTCAAGGTCAAGGACGAGGAGGCA 1069
QY 1021 TCGGGACGAGCGGAGCGGCGGAGGATGCTTACAGCTTGGCTGCGCTCGCGAGTTT 1080
DB 1070 TT-----CTCGCGAGGACCGCTTACCCCTTTCGCGAGCTCTCTCAGTGGC 1114
QY 1081 TCGGGGCGGCTTTCGACAGCTCGCATGCGATGACCGGGTGTCTGACGATCGAGCTGAACG 1140
DB 1115 TCGGCGCGCTCGTCAGCGACTCTATTACGCGCCACGCGCTCTCACCATCGAGCGCGGCC 1174
QY 1141 CGGTGACGCAATTCGGTGTTCGCGCCGATGGGAGCGTGCCTCGCGCTCGACCGGGGCA 1200
DB 1175 AGTCGACGACCGCAACCCCTCTCATCGACTCGAGAACAGAGTTCGCAACACGCGGCA 1234
QY 1201 ATTTCATGGGCCAGCATGTGGCGCTGACGTCCGATGCTCGCGACCGCGCTCACCCTTTC 1260
DB 1235 ATTTCAGGCTCGGCTGTGGCCACACCATGAGAGAGTCTCGCTCGGCTCGCCGAGA 1294
QY 1261 TGGCGGCGCTTTCGGGAGCGCAGATTGCACTGTGACAGATGAAGGCTGAACCGTGGGC 1320
DB 1295 TCGGCAAGCTCAACTTCAACGAGCTCACCGAGATGCTCAACCGCGGATGAACCGCGGCC 1354
QY 1321 TGGCCCGCTTCTCCACCGGGGCGCGCGGGTGAATTCGGGCTTCAATGGGCGGACAGG 1380
DB 1355 TCCCTCTCTGCTCGGCGGCAAGAGCCCTCTCTCTCTACCACTGCAAGGGCTCTGACA 1414

QY 1381 TGACGGCGACCGCGCTCTCTGGCGGAGATCGCA---GCCACGGGACCTGCTCGATCCATT 1437
DB 1415 TCGCGCTCGCGGTACACCTCGGAGTTGGAGACACCTCGCAACCTGTGACGACGATG 1474
QY 1438 CGATCTTCAGGAACGCGCGCAATCAGGATGTGGTCTCGCTTGGGACCATCGCGCGGCC 1497
DB 1475 TCCAGCGGCTGAGATGGGGAACACAGCGGTCAACTCGCTTGGCTCATCTCGGCTCGTC 1534
QY 1498 TCTGCGCGCAGAAGATCGACCGTTGGCGGAGATCTTGGGATCTCTCGCTCTCTGCTTTG 1557
DB 1535 GCACGACCGAGTCCACAGACGCTCTTCTCTCTCGCCACCCACCTCTACTGCGTTC 1594
QY 1558 CACAAGCTCGGAGCTGCGC 1577
DB 1595 TCCAAGCCATCGACTTGGCG 1614

RESULT 15

US-09-765-873A-7
; Sequence 7, Application US/09765873A
; Patent No. 6521748
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BC1009 US CIP
; CURRENT APPLICATION NUMBER: US/09/765,873A
; CURRENT FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 7
; LENGTH: 2151
; TYPE: DNA
; ORGANISM: Rhodotorula glutinis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2151)
US-09-765-873A-7

Query Match 5.6%; Score 103.2; DB 3; Length 2151;
Best Local Similarity 46.5%; Pred. No. 1.5e-11;
Matches 511; Conservative 0; Mismatches 553; Indels 36; Gaps 4;
QY 499 CTGAGGAGCTCCGGTGCAGGAGGGACCATCGCTCGCTGATCGACCTGCTCAATT 558
DB 530 CCGCGGCACTGGCTGTCCGCTCGCTCGAGGGCTCACCACATTTCTCAACC 589
QY 559 CCGAGCTCGCTCCGGCGGTTCCAGCGCGGACCGTGGCGGCTCGGGTGACCTGACAC 618
DB 590 ACGGCATCACCCCATCGTCCCTCCGGCGCACCATCTCTCGCGGACCATCTCTGCTCGGGGACCTGTCTC 649
QY 619 CGCTTGGCATATGCTCTGCTCTCAGGCGCGGGAGACT-----TCC 663
DB 650 CTCTCTCTACATTCGAGCGGCAATCAGCGGTCAACCGGACGAGGTCAGCTGCTC 709
QY 664 TGGACCGGACCGGACCGCGCTTTCGAGCGGCAAGGAGTTCGCGCGCGGACCGCTGCAAC 723
DB 710 ACGAGGCAAGGAGAGATCTCTGTAACCGCGGAGCGATGGCGCTCTTCAACCTCGAGC 769
QY 724 CGCTCGATCTCTCCCATCGGATGCACTGCGCTGTCACCGGACCTCGGACCTCGCCATGACCG 783
DB 770 CCGTCTGCTCTCGGCGGAGGAGGTCTCGGTCTCGTCAACGGCACCGCGCTCTAGCAT 829
QY 784 GGATCGCGCTGTGAATGCTCAGCGCTCGCGCATCTCGGCAACTGGGCGGTGGCGTTGA 843
DB 830 CGATGGCCACCTCGCTCTGACGACGCTCACAATGCTCTCGCTCTCTCGAGTCTGCTCA 889
QY 844 CGGCTCTGCTTGGGAGATGTCTGAGAGCGGACCGAGGATGGGCGGCGGACCTGTCG 903
DB 890 CGGCAATGAGCTGCGGCGGAGAGTGGTTCGCGCTCAGCGCGGCTCTTCCACCCCTTCTTCAACG 949

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 10, 2005, 06:02:09 ; Search time 1553 Seconds
(without alignments)
9850.837 Million cell updates/sec

Title: US-10-621-826-2

Perfect score: 1850

Sequence: 1 ccgagggcgccatccacgag.....tcagctcgatccgcagcagg 1850

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134589005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA Main.*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB	ID	Description
1	1850	100.0	1850	7	US-10-621-826-2	Sequence 2, Appli
2	1580	85.4	1581	10	US-11-069-633-7	Sequence 7, Appli
3	1446	78.2	1446	6	US-10-369-493-31503	Sequence 31503, A
4	561	20.3	1626	7	US-10-464-609-3	Sequence 3, Appli
5	244.2	13.2	1587	6	US-10-369-493-32499	Sequence 32499, A
c 6	242.4	13.1	42000	7	US-10-159-257A-1	Sequence 1, Appli
c 7	242.4	13.1	63158	6	US-10-292-198-1	Sequence 1, Appli
8	238.4	12.9	1497	6	US-10-369-493-31670	Sequence 31670, A
9	232.8	12.6	1542	6	US-10-156-761-3312	Sequence 3312, Ap
c 10	232.8	12.6	9025608	6	US-10-156-761-1	Sequence 1, Appli
11	231.6	12.5	1521	7	US-10-282-122A-13842	Sequence 13842, A
12	215.4	11.6	1503	6	US-10-369-493-42898	Sequence 42898, A
13	211.6	11.4	1503	6	US-10-369-493-28443	Sequence 28443, A
14	211.6	11.4	1503	6	US-10-369-493-31202	Sequence 31202, A
15	211.6	11.4	1521	7	US-10-282-122A-13653	Sequence 13653, A
16	177.4	9.6	1518	6	US-10-369-493-28430	Sequence 28430, A
17	177.4	9.6	1518	6	US-10-369-493-31190	Sequence 31190, A
18	177.2	9.6	1575	6	US-10-369-493-47139	Sequence 47139, A
19	173.4	9.4	1530	3	US-09-815-242-7987	Sequence 7987, Ap
20	173.4	9.4	1530	6	US-10-369-493-40527	Sequence 40527, A
21	173.4	9.4	1530	7	US-10-282-122A-30600	Sequence 30600, A
22	172.8	9.3	2256646	7	US-10-470-565-1	Sequence 1, Appli
23	170.8	9.2	1530	6	US-10-369-493-35187	Sequence 35187, A

24	170.8	9.2	1533	6	US-10-369-493-38162	Sequence 38162, A
25	170.8	9.2	1533	6	US-10-369-493-38793	Sequence 38793, A
26	170.8	9.2	1542	6	US-10-369-493-36026	Sequence 36026, A
27	161.8	8.7	1542	3	US-09-833-745-7	Sequence 7, Appli
28	161.6	8.7	1521	6	US-10-369-493-39101	Sequence 39101, A
29	161.6	8.7	1521	6	US-10-369-493-39469	Sequence 39469, A
30	161.6	8.7	1521	6	US-10-369-493-39852	Sequence 39852, A
31	159.4	8.6	1566	6	US-10-369-493-32807	Sequence 32807, A
32	157	8.5	1525	7	US-10-282-122A-23382	Sequence 23382, A
33	155.2	8.4	1545	7	US-10-282-122A-33163	Sequence 33163, A
34	153.8	8.3	1536	3	US-09-833-745-12	Sequence 12, Appli
35	151.8	8.2	1446	6	US-10-369-493-28511	Sequence 28511, A
36	151.8	8.2	1491	6	US-10-369-493-31270	Sequence 31270, A
37	151.6	8.2	1452	6	US-10-369-493-32328	Sequence 32328, A
38	149.4	8.1	1530	6	US-10-369-493-37377	Sequence 37377, A
39	146.6	7.9	1530	7	US-10-282-122A-31973	Sequence 31973, A
40	144.6	7.8	1461	6	US-10-369-493-27975	Sequence 27975, A
41	144.6	7.8	1506	6	US-10-369-493-30731	Sequence 30731, A
42	144.6	7.8	1533	7	US-10-282-122A-12902	Sequence 12902, A
43	141.2	7.6	1362	6	US-10-369-493-37343	Sequence 37343, A
44	140.8	7.6	1386	6	US-10-369-493-42300	Sequence 42300, A
45	133.4	7.2	1521	6	US-10-369-493-32286	Sequence 32286, A

ALIGNMENTS

RESULT 1

US-10-621-826-2
; Sequence 2, Application US/10621826
; Publication No. US20040059103A1
; GENERAL INFORMATION:
; APPLICANT: E.I. du Pont de Nemours and Company, Inc.
; APPLICANT: Huang, Lixuan
; APPLICANT: Xue, Zhixiong
; TITLE OF INVENTION: DNA and Amino Acid Sequences of a Tyrosine Ammonia Lyase Enzyme
; TITLE OF INVENTION: From the Bacterium Rhodobacter sphaeroides
; FILE REFERENCE: CL2169 US NA
; CURRENT APPLICATION NUMBER: US/10/621.826
; CURRENT FILING DATE: 2003-07-17
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 1850
; TYPE: DNA
; ORGANISM: R. sphaeroides
US-10-621-826-2

Query Match	100.0%	Score 1850;	DB 7;	Length 1850;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1850;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	CCGAGGCGGCCCATCCACGAGATGCGCGCCCGTTTCGGCCCTGACGCCCGCGGATCTCTCGC	60	
Db	1	CCGAGGCGGCCCATCCACGAGATGCGCGCCCGTTTCGGCCCTGACGCCCGCGGATCTCTCGC	60	
Qy	61	TCGATCTCTGGGCCCTCTGGCGGCGACGATCGAAGCAAGAACTCTGCATTAACCAAGCTTC	120	
Db	61	TCGATCTCTGGGCCCTCTGGCGGCGACGATCGAAGCAAGAACTCTGCATTAACCAAGCTTC	120	
Qy	121	CACCTACCTTGGCGAAGACAGAGGAGTGAAGCAATGCTCGCCATGAGCCCCCGGAGC	180	
Db	121	CACCTACCTTGGCGAAGACAGAGGAGTGAAGCAATGCTCGCCATGAGCCCCCGGAGC	180	
Qy	181	CGGCCGTCGAGCTGGATCGCCACATCGATCTGGACACAGGCCCATGCGTGGCGAGCGCG	240	
Db	181	CGGCCGTCGAGCTGGATCGCCACATCGATCTGGACACAGGCCCATGCGTGGCGAGCGCG	240	
Qy	241	CGCGCGCGGATTTGCTTTGCCCTTCCGGCGCGCAGCCGGTGCCTGCGTCCGAAGCGCGC	300	
Db	241	CGCGCGCGGATTTGCTTTGCCCTTCCGGCGCGCAGCCGGTGCCTGCGTCCGAAGCGCGC	300	
Qy	301	TCGGCGCTGTCTATCCGCGAGCGGCCCATGTCTTAGGACTGACAAACCGGCTTCGGTCCCC	360	

Db 301 |||||TCGGCGTGTATCCGCGAGCGCGCATGTCTACGGACTGACAAACCGGCTTCGGTCCCC360
Qy 361 |||||TTGCGAACCGCTTGATCTCAGGTGAGAAATGTCGGAACGCTGCGGCGCAATCTTGTCCATC420
Db 361 |||||TTGCGAACCGCTTGATCTCAGGTGAGAAATGTCGGAACGCTGCGGCGCAATCTTGTCCATC420
Qy 421 |||||ATCTGGCCAGCGGCGTGGGACCGGTGCTTGACTGGAACGCGGCGCGCCCATGTTCTGG480
Db 421 |||||ATCTGGCCAGCGGCGTGGGACCGGTGCTTGACTGGAACGCGGCGCGCCCATGTTCTGG480
Qy 481 |||||CGGCTCTGTTGATCGCTCAGGAGCCTTCCGCTGCGAGCGGAGCGGACCATCGCTCGCC540
Db 481 |||||CGGCTCTGTTGATCGCTCAGGAGCCTTCCGCTGCGAGCGGAGCGGACCATCGCTCGCC540
Qy 541 |||||TGATCGACCTGCTCAATTCGAGACTCGCTCCGCGCGTTCACGCGCGGACCGTGGGG600
Db 541 |||||TGATCGACCTGCTCAATTCGAGACTCGCTCCGCGCGTTCACGCGCGGACCGTGGGG600
Qy 601 |||||CGTGGGTGACCTGACACCGCTTTCGCGCATATGTTGCTTCTGCTCCAGGGCGCGGAGACT660
Db 601 |||||CGTGGGTGACCTGACACCGCTTTCGCGCATATGTTGCTTCTGCTCCAGGGCGCGGAGACT660
Qy 661 |||||TCCTGGACCGGAGCGGAGCGGCTTGACCGGCGAGAAAGGCTTCGCGCGGAGCGGCTGC720
Db 661 |||||TCCTGGACCGGAGCGGAGCGGCTTGACCGGCGAGAAAGGCTTCGCGCGGAGCGGCTGC720
Qy 721 |||||AACCGCTCGATCTCTCCCATCGGAGTCACTGCGGCTGTTCAACGGAGCCTCGCCCATGA780
Db 721 |||||AACCGCTCGATCTCTCCCATCGGAGTCACTGCGGCTGTTCAACGGAGCCTCGCCCATGA780
Qy 781 |||||CCGGGATCGCGCTGGTGAATGCTCACGCCCTGCGCCCATCTCGGCAACTGCGGCGTGGCGT840
Db 781 |||||CCGGGATCGCGCTGGTGAATGCTCACGCCCTGCGCCCATCTCGGCAACTGCGGCGTGGCGT840
Qy 841 |||||TGAACGCGCTTCTGCGGAATGCTGAGAGCGCGGACCGGAGGATGCGGCGCGGCACTGT900
Db 841 |||||TGAACGCGCTTCTGCGGAATGCTGAGAGCGCGGACCGGAGGATGCGGCGCGGCACTGT900
Qy 901 |||||CCGACCTGCGCGCATCCCGGACAGAGACCGCGGAGGCTTCGCGCGCGGCTGCGGCGTGG960
Db 901 |||||CCGACCTGCGCGCATCCCGGACAGAGACCGCGGAGGCTTCGCGCGCGGCTGCGGCGTGG960
Qy 961 |||||ACGGCAGCGCGCGGTGCTCGGCACTGTCGAGCGGAGGCTTCGAGCGCGGCACTGTA1020
Db 961 |||||ACGGCAGCGCGCGGTGCTCGGCACTGTCGAGCGGAGGCTTCGAGCGCGGCACTGTA1020
Qy 1021 |||||TCGGGACGGAGCGGAGCGGCGAGATGCTTACAGCTTGCCTGCGCTTCGCGAGGTTTC1080
Db 1021 |||||TCGGGACGGAGCGGAGCGGCGAGATGCTTACAGCTTGCCTGCGCTTCGCGAGGTTTC1080
Qy 1081 |||||TCGGGCGGGCTTCGACAGCTCGCATGGCATGACCGGGTGTGACGATCGAGCTGAACG1140
Db 1081 |||||TCGGGCGGGCTTCGACAGCTCGCATGGCATGACCGGGTGTGACGATCGAGCTGAACG1140
Qy 1141 |||||CGGTGACCGGACAAATCCGGTGTTCGCGCCGATGGCAGCGTCCCGCTTCGACGGGGCA1200
Db 1141 |||||CGGTGACCGGACAAATCCGGTGTTCGCGCCGATGGCAGCGTCCCGCTTCGACGGGGCA1200
Qy 1201 |||||ATTTCATGGGCCAGCATGTCGCGCTGACGTCGATGCGCTTCGCCACGCGCGTCAACGTTTC1260
Db 1201 |||||ATTTCATGGGCCAGCATGTCGCGCTGACGTCGATGCGCTTCGCCACGCGCGTCAACGTTTC1260
Qy 1261 |||||TGGCGGCTTCGAGAGCGGAGATTCAGCTCTGACAGTGAAGGCTGMAACGCTGGGC1320
Db 1261 |||||TGGCGGCTTCGAGAGCGGAGATTCAGCTCTGACAGTGAAGGCTGMAACGCTGGGC1320
Qy 1321 |||||TGCCCCCTTCCTCCACCGGGGCGCGCGGGTGAATTCGGCTTCATGGGGCGACAGG1380
Db 1321 |||||TGCCCCCTTCCTCCACCGGGGCGCGCGGGTGAATTCGGCTTCATGGGGCGACAGG1380
Qy 1381 |||||TGAACGCGACTGCTGCGCGAGATGCGGACCGGACCGGACTGCTCGATCCATTCGA1440

Db 1381 |||||TGACGCGGACCGCGCTCTCTGGCGAGATGCGAGCCACGGGACCTGCCTCGATCCATTCGA1440
Qy 1441 |||||TCTCCACGAAACCGCGCAATCAGGATGTGTCTCGCTTCGGACCATCGCGCGCGCTCT1500
Db 1441 |||||TCTCCACGAAACCGCGCAATCAGGATGTGTCTCGCTTCGGACCATCGCGCGCGCTCT1500
Qy 1501 |||||GCCGCGAGAAGATCGACCGTTCGGCGGAGATCCTTCGATTCCTCTGTCTTTCGAC1560
Db 1501 |||||GCCGCGAGAAGATCGACCGTTCGGCGGAGATCCTTCGATTCCTCTGTCTTTCGAC1560
Qy 1561 |||||AAGTCGCGAGCTGCGCTCGCGGAGCGGCTAGACGGGGTGTCTCCGCGGGGAAAGC1620
Db 1561 |||||AAGTCGCGAGCTGCGCTCGCGGAGCGGCTAGACGGGGTGTCTCCGCGGGGAAAGC1620
Qy 1621 |||||TGTGTCAGCGCTTCGCGGAGCGGCTTCGCGGCTTCGAGACGACCGGCGCTTCGGACAGG1680
Db 1621 |||||TGTGTCAGCGCTTCGCGGAGCGGCTTCGCGGCTTCGAGACGACCGGCGCTTCGGACAGG1680
Qy 1681 |||||AAATTGCGCGCTTGTCTACGCACTCTTTCGCAATCTCCCGTTCGAGCGCGGCTCAGG1740
Db 1681 |||||AAATTGCGCGCTTGTCTACGCACTCTTTCGCAATCTCCCGTTCGAGCGCGGCTCAGG1740
Qy 1741 |||||TGTGCGGGGATCCACCGTTCGAGGCTTCGAGCTTCGCAATTCGAGCGAGTTCGATCG1800
Db 1741 |||||TGTGCGGGGATCCACCGTTCGAGGCTTCGAGCTTCGCAATTCGAGCGAGTTCGATCG1800
Qy 1801 |||||ACGCCACGACGCGCAGGACTTCGATGGCGATCAGCTTCGATCCGACGAGG1850
Db 1801 |||||ACGCCACGACGCGCAGGACTTCGATGGCGATCAGCTTCGATCCGACGAGG1850

RESULT 2
US-11-069-633-7
; Sequence 7, Application US/11069633
; Publication No. US20050208643A1
; GENERAL INFORMATION:
; APPLICANT: Schmidt-Dannert, Claudia
; APPLICANT: Watte, Kevin
; TITLE OF INVENTION: Flavonoids
; FILE REFERENCE: 09531-160001
; CURRENT APPLICATION NUMBER: US/11/069,633
; CURRENT FILING DATE: 2005-03-01
; PRIOR APPLICATION NUMBER: US 60/549,077
; PRIOR FILING DATE: 2004-03-01
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1581
; TYPE: DNA
; ORGANISM: Rhodobacter sphaeroides
US-11-069-633-7

Query Match 85.4%; Score 1580; DB 10; Length 1581;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1580; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 148 TGAAGCCAAATGCTCGCCATGAGCCCCCGAAGCCGSCGCTCGAGCTGGATCGCCACATCG 207
Db 2 TGAAGCCAAATGCTCGCCATGAGCCCCCGAAGCCGSCGCTCGAGCTGGATCGCCACATCG 61
Qy 208 ATCTGGACAGGCCCATATGCTGGCGAGCGGCGCGGATGTCTTGGCCCTCCGG 267
Db 62 ATCTGGACAGGCCCATATGCTGGCGAGCGCGCGGATGTCTTGGCCCTCCGG 121
Qy 268 CGCGGACCGGTCGCTGCTGCGAAGCGCGGCTGCGGCTGTCTTCGCGGCGCGCC 327
Db 122 CGCGGACCGGTCGCTGCTGCGAAGCGCGGCTGCGGCTGTCTTCGCGGCGCGCC 181
Qy 328 ATGTCTACGAGCTGACAAACCGCTTCGGTCCCTTCGGAACCGCCTGATCTCAGGTGAGA 387
Db 182 ATGTCTACGAGCTGACAAACCGCTTCGGTCCCTTCGGAACCGCCTGATCTCAGGTGAGA 241
Qy 388 ATGTCCGAAACGCTGACGGCCAATCTTGTCCATCATCTGGCCAGCGGCTGGGACCGGTGC 447


```

Db      242  ATGTCGGAACGCTGCAGGCCAATCTTGTCCATCATCTGGCCAGCGCGTGGGACCGGTGC 301
QY      448  TTGACTGAGACGACGGCGCGCCATAGTGTTCGGCGCGTGTGTGTCGATCGCTCAGGAG 507
Db      302  TTGACTGAGACGACGGCGCGCCATAGTGTTCGGCGCGTGTGTGTCGATCGCTCAGGAG 361
QY      508  CCTCCGGTCCAGCGAGGGACCAATCGCTCGCCTGATCGACTGCTCAATTCGAGCTCG 567
Db      362  CCTCCGGTCCAGCGAGGGACCAATCGCTCGCCTGATCGACTGCTCAATTCGAGCTCG 421
QY      568  CTCGGCGCTTCCACGCGCGCGCACGGTGGGCGCGTGGGCTGACCTGACACCGCTTGGC 627
Db      422  CTCGGCGCTTCCACGCGCGCGCACGGTGGGCGCGTGGGCTGACCTGACACCGCTTGGC 481
QY      628  ATATGGTGTCTGTCTCAGGGCCGGGAGACTTCTCTGGACCGGGACCGGACCGGCTTG 687
Db      482  ATATGGTGTCTGTCTCAGGGCCGGGAGACTTCTCTGGACCGGGACCGGACCGGCTTG 541
QY      688  ACGGCGAGAGGGCTCCGGCGCGGACCGCTGCAACCGCTCGATCTCTCCATCGCGATG 747
Db      542  ACGGCGAGAGGGCTCCGGCGCGGACCGCTGCAACCGCTCGATCTCTCCATCGCGATG 601
QY      748  CACTGGCGCTGTCAACGGGACCTCCGCGATGACCGGATCGCGTGTGTAATGCTCACG 807
Db      602  CACTGGCGCTGTCAACGGGACCTCCGCGATGACCGGATCGCGTGTGTAATGCTCACG 661
QY      808  CCTGCCGCCATCTCGGCAACTGGGCGGTGGGCTTGAACGGCCCTGCTTGGCGAATGCTCGA 867
Db      662  CCTGCCGCCATCTCGGCAACTGGGCGGTGGGCTTGAACGGCCCTGCTTGGCGAATGCTCGA 721
QY      868  GAGGCGGACCGAGCATGGGCGCGGACACTGTGCGACCTCGGCGCGATCCCGACAGA 927
Db      722  GAGGCGGACCGAGCATGGGCGCGGACACTGTGCGACCTCGGCGCGATCCCGGACAGA 781
QY      928  AGGACGCGCAGCAGGCTGCGCGCCGCGTGGACGGCAGCGCGGCTGTTCGGGACG 987
Db      782  AGGACGCGCAGCAGGCTGCGCGCCGCGTGGACGGCAGCGCGGCTGTTCGGGACG 841
QY      988  TCATTGGCGAGCGAGGCTCGACCGCGCGCATATCGGACCGGACCGGACCGGCGGCGAGG 1047
Db      842  TCATTGGCGAGCGAGGCTCGACCGCGCGCATATCGGACCGGACCGGACCGGCGGCGAGG 901
QY      1048  ATGCTTACAGCTCGCTGCGCTCGGCGAGTCTTCGGGGCGGGCTTGACACGCTCGAT 1107
Db      902  ATGCTTACAGCTCGCTGCGCTCGGCGAGTCTTCGGGGCGGGCTTGACACGCTCGCAT 961
QY      1108  GGCATGACCGGGTGTGACGATCGAGCTGAACGGGTGACCGGCTGACCAATCCGGTGTTCGCG 1167
Db      962  GGCATGACCGGGTGTGACGATCGAGCTGAACGGGTGACCGGCTGACCAATCCGGTGTTCGCG 1021
QY      1168  CCGATGGCAGCGTCCCGCCCTGACCGGGGCAATTTTCATGGGCGAGCATGTGGCGCTGA 1227
Db      1022  CCGATGGCAGCGTCCCGCCCTGACCGGGGCAATTTTCATGGGCGAGCATGTGGCGCTGA 1081
QY      1228  CGTCCGATCGCTCGCACGCGCGTCAACGTTTCGGGGGCGCTTCGGGAGCGCCAGATTG 1287
Db      1082  CGTCCGATCGCTCGCACGCGCGTCAACGTTTCGGGGGCGCTTCGGGAGCGCCAGATTG 1141
QY      1288  CACCTCTGACAGATGAAGGCTGAACCGTGGGCTGCCCGCTTCTCCACCGGGGCGCGG 1347
Db      1142  CACCTCTGACAGATGAAGGCTGAACCGTGGGCTGCCCGCTTCTCCACCGGGGCGCGG 1201
QY      1348  CCGGGTTGAATTCGGGTTTCATGGGCGCACAGGTGAAGGGACCGCGCTCTCTGGCCGAGA 1407
Db      1202  CCGGGTTGAATTCGGGTTTCATGGGCGCACAGGTGAAGGGACCGCGCTCTCTGGCCGAGA 1261
QY      1408  TGCAGGCCACGGGACCTGCTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 1467
Db      1262  TGCAGGCCACGGGACCTGCTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 1321
QY      1468  TGCTCTCGCTTGGGACCATCGCGCGCGCTCTCGCGGAGAGATCGACCGCTTGGCGG 1527

```

```

Db      1322 TGCTCTCGCTTGGGACCATCGCGCGCGCTCTGCGCGAGAGATCGACCGTTGGCGG 1381
QY      1528 AGATCTTTGCGATCTCGCTCTCTGTCTTGTGCAAGCTGCGGAGCTGCGTGCAGCAGC 1587
Db      1382 AGATCTTTGCGATCTCGCTCTCTGTCTTGTGCAAGCTGCGGAGCTGCGTGCAGCAGC 1441
QY      1588 GCCTAGACGGGGTGTCTCCCGGGGAGAGAGCTGTGTCAGGCCCTGCGGAGCAGTTCC 1647
Db      1442 GCCTAGACGGGGTGTCTCCCGGGGAGAGAGCTGTGTCAGGCCCTGCGGAGCAGTTCC 1501
QY      1648 CGCGCTTTGACACGACCGCGCCCTTGGGACAGGAAATTCGCGCGCTTGTCTACGACCTCT 1707
Db      1502 CGCGCTTTGACACGACCGCGCCCTTGGGACAGGAAATTCGCGCGCTTGTCTACGACCTCT 1561
QY      1708 TGCAGCAATCTCCCGTCTGA 1727
Db      1562 TGCAGCAATCTCCCGTCTGA 1581

RESULT 3
US-10-369-493-31503
; Sequence 31503, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 31503
; LENGTH: 1446
; TYPE: DNA
; ORGANISM: Rhodobacter sphaeroides
US-10-369-493-31503

Query Match      78.2%; Score 1446; DB 6; Length 1446;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1446; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      273  GACCGGTGCGGTGCGTCCGAGCGCGCTCGGCGCTGTCTATCCGAGGCGCGCATGTC 332
Db      1  GACCGGTGCGGTGCGTCCGAGCGCGCTCGGCGCTGTCTATCCGAGGCGCGCATGTC 60
QY      333  TACCGACTGACAAACCGGCTTCGGTCCCTTTCGGAACCGCCCTGATCTCAGGTGAGAATGTC 392
Db      61  TACCGACTGACAAACCGGCTTCGGTCCCTTTCGGAACCGCCCTGATCTCAGGTGAGAATGTC 120
QY      393  CGAACGCTGAGGCAATCTTGTCCATCATCTGCGCAGCGCGTGGGACCGGTGTTGAC 452
Db      121  CGAACGCTGAGGCAATCTTGTCCATCATCTGCGCAGCGCGTGGGACCGGTGTTGAC 180
QY      453  TGGACGAGCGCGCGCATGTTCTTGGCGGTCTGCTGATCGCTCAGGAGCGCTCC 512
Db      181  TGGACGAGCGCGCGCATGTTCTTGGCGGTCTGCTGATCGCTCAGGAGCGCTCC 240
QY      513  GGTGCCAGCGAGGGGACCATCGCTCGCTCGATCGACCTGTCTCAATTCGAGCTCGTCCG 572
Db      241  GGTGCCAGCGAGGGGACCATCGCTCGCTCGATCGACCTGTCTCAATTCGAGCTCGTCCG 300
QY      573  GCGCTTCCGAGCGCGGACCGGTGGGCGCGTGGGTGACCTGACACCGCTTGGCATATG 632
Db      301  GCGCTTCCGAGCGCGGACCGGTGGGCGCGTGGGTGACCTGACACCGCTTGGCATATG 360
QY      633  GTGCTCTGCTTCCAGGGCGGGGAGACTTCTTGGACCGGGACGGGACCGCTTGGCGG 692

```

Db 361 GTGCTCTGCTCCAGAGGCGCGGAGAGCTTCTGTGACCGGGACGGGACGCGGCTTTGACGGC 420
Qy GCAGAGGGGCTCCGGCGGGAGCGCTCAACCGCTCGATCTCTCCCATCGGATGCACTG 752
Db 421 GCAGAGGGGCTCCGGCGGGAGCGCTGCAACCGCTCGATCTCTCCCATCGGATGCACTG 480
Qy 753 GCGCTGCTCAACGGGAGCTCCGCGCATGACCGGGATCGCGTGTGTGAATGTCTACGCCCTGC 812
Db 481 GCGCTGCTCAACGGGAGCTCCGCGCATGACCGGGATCGCGTGTGTGAATGTCTACGCCCTGC 540
Qy 813 GCGCATCTCGGCAACTGGCGGTGGCGTTGACGGCCCTGCTTGGCGGAATGTCTGAGAGGC 872
Db 541 CGCCATCTCGGCAACTGGCGGTGGCGTTGACGGCCCTGCTTGGCGGAATGTCTGAGAGGC 600
Qy 873 CGGACCGAGGATGGGCGGGGACTCTCGGACCTGGCGGCGCATCCGGACGAGAGGC 932
Db 601 CGGACCGAGGATGGGCGGGGACTCTCGGACCTGGCGGCGCATCCGGACGAGAGGC 660
Qy 933 GCGGACCGAGGCTCGCGCCCGCGTGGACGGCGCGCGGGTGTCTCCGGCAGCTCAT 992
Db 661 GCGGACCGAGGCTCGCGCCCGCGTGGACGGCGCGGGTGTCTCCGGCAGCTCAT 720
Qy 993 GCGGACCGAGGCTCGAGCGCGCGGATATCGGAGCGGAGCGGGGCGAGGATGCC 1052
Db 721 GCGGACCGAGGCTCGAGCGCGCGGATATCGGAGCGGAGCGGGGCGAGGATGCC 780
Qy 1053 TACAGCGCTGCGCTCGCGAGGTTCTCGGGGCGGGCTTCGACACGCTCGCATGGCAT 1112
Db 781 TACAGCGCTGCGCTCGCGAGGTTCTCGGGGCGGGCTTCGACACGCTCGCATGGCAT 840
Qy 1113 GACCGGCTGTGACGATCGAGCTGAACCGCGGTGACCGCAATCCGGTGTCTTCGCGCCGAT 1172
Db 841 GACCGGCTGTGACGATCGAGCTGAACCGCGGTGACCGCAATCCGGTGTCTTCGCGCCGAT 900
Qy 1173 GCGAGCGTCCCGCTCGCGAGGCAATTCATGGGCGCAATTCATGGGCGGAGTGGCGTTCG 1232
Db 901 GCGAGCGTCCCGCTCGCGAGGCAATTCATGGGCGGCAATTCATGGGCGGAGTGGCGTTCG 960
Qy 1233 GATGCGCTCGCGAGGCGCTCACGTTCTCGCGGCGCTTCGGGAGCGGAGTGGCGTTCGAGCT 1292
Db 961 GATGCGCTCGCGAGGCGCTCACGTTCTCGCGGCGCTTCGGGAGCGGAGTGGCGTTCGAGCT 1020
Qy 1293 CTGACAGATGAAGGCTGAACCGTGGGCTGCGCCCTCTCTCCACCGGGGCGCCCGCGGG 1352
Db 1021 CTGACAGATGAAGGCTGAACCGTGGGCTGCGCCCTCTCTCCACCGGGGCGCCCGCGGG 1080
Qy 1353 TTGAATTCGGGCTTCATGGGCGCACAGGTGACGGCGACCGCGCTCTCTGGCGGAGATGCGA 1412
Db 1081 TTGAATTCGGGCTTCATGGGCGCACAGGTGACGGCGACCGCGCTCTCTGGCGGAGATGCGA 1140
Qy 1413 GCCACGGGAGCTGCTCGATCCATTGATCTCCACGAAACCGCCCATCAGGATGTGTC 1472
Db 1141 GCCACGGGAGCTGCTCGATCCATTGATCTCCACGAAACCGCCCATCAGGATGTGTC 1200
Qy 1473 TCGCTTGGGACCATCGCGCGGCTCTGCGGAGAGATTCGACCCCTTGGGCGGAGATC 1532
Db 1201 TCGCTTGGGACCATCGCGCGGCTCTGCGGAGAGATTCGACCCCTTGGGCGGAGATC 1260
Qy 1533 CTTGCGATCTCGTCTCTGTTCGACAGCTGCGGAGCTGCGCTGCGGAGCGGCGCTA 1592
Db 1261 CTTGCGATCTCGTCTCTGTTCGACAGCTGCGGAGCTGCGCTGCGGAGCGGCGCTA 1320
Qy 1593 GACGGGCTGCTCCCGGGGAGAGCTGTCAGGCGCCCTGCGGAGAGTTCGCGCG 1652
Db 1321 GACGGGCTGCTCCCGGGGAGAGCTGTCAGGCGCCCTGCGGAGAGTTCGCGCG 1380
Qy 1653 CTTGAGACGAGCGGCGCTTCGGACAGAAATTCGCGCTTGTGACGACCTCTTTCGAG 1712
Db 1381 CTTGAGACGAGCGGCGCTTCGGACAGAAATTCGCGCTTGTGACGACCTCTTTCGAG 1712
Qy 1713 CAATCT 1718
Db 1441 CAATCT 1446

RESULT 4

US-10-464-609-3
; Sequence 3, Application US/10464609
; Publication No. US20040029230A1
; GENERAL INFORMATION:
; APPLICANT: KYNDT, John, Jozef Armand
; APPLICANT: VAN BEEUMEN, Jozef
; TITLE OF INVENTION: No. US20040029230A1el Methods For Synthesis of
; FILE OF INVENTION: Holo-Photoactive Yellow Protein
; FILE REFERENCE: 50304/008001
; CURRENT APPLICATION NUMBER: US/10/464,609
; CURRENT FILING DATE: 2003-06-18
; PRIOR APPLICATION NUMBER: US 60/389,593
; PRIOR FILING DATE: 2002-06-18
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1626
; TYPE: DNA
; ORGANISM: Rhodobacter capsulatus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1623)
; OTHER INFORMATION: n = A,T,C or G

Query Match 30.3%; Score 561; DB 7; Length 1626;
Best Local Similarity 60.9%; Pred. No. 2.6e-137;
Matches 947; Conservative 0; Mismatches 603; Indels 6; Gaps 2;
Qy 175 CGAAGCGGCGCGGATGCTGCTTTCGCCCTCCGCGCGGCGACCGGTCGCTGCCAAG 234
Db 53 CCAAGACTGCTTCGCGCTGGACGGGGCGCTGACACTTGTCCAATGAGGCGCATCGCA 112
Qy 235 GCGGCGCGCGCGGATGCTGCTTTCGCCCTCCGCGCGGCGACCGGTCGCTGCCAAG 294
Db 113 CACATCGCAGCCGATTTCCGTGACCCCGCGCTGCGGAGCGCTGCGCGGGCCCATG 172
Qy 295 GCGGCTTCGCGCTGTGTCATCCGCGAGGCGCGCATGTCTACGAGCTGACAAACCGGCTTCG 354
Db 173 CCGGCTTTCGAGCAGCGCATCGCGGAGCGCGCACATTTACGGCATCACACCGGCTTCG 232
Qy 355 GTCCTCTTCGAAACCGCTGATCTCAGGTGAGAAATGTCGAAACGCTGCGGCAATCTTG 414
Db 233 GCGGCTTCGCGAAACCGCTGATCGGGGCGGATCAGGGGCGGAGCTGACAGAACCTGA 292
Qy 415 TCCATCATCTGGCGAGCGGCTGGGACCGGTGCTTGAATGAGACGAGCGCGCGCCATGG 474
Db 293 TCTATCATCTGGCGACCGGCTCGGGCCGAAACTGAGCTGGGCGGAGCGCGGGCTTGA 352
Qy 475 TTTGCGGCGCTTCGTGTGATCGCTCAGGAGCGCTCCCGGTGCGAGGAGGACCATCG 534
Db 353 TGTGCGGCGGCTCAACTCGATCTCTCAAGGCGCGCTCGGGGCGCTCGCCGAGACGATCG 412
Qy 535 CTCGCTGATTCGACCTGCTCAATTCGAGCTCGCTCCGCGCGTTCGCGCGGCGCACCG 594
Db 413 ACCGATTCGTTCGCTGCTCAATTCGCGGGTTCGCGCGGCTTCGCGCGAGGAAACCG 472
Qy 595 TGGGCGGCTCGGCTGACCTGACACCGCTTCGCGCATATGCTGCTTCGCTCCAGGCGCGG 654
Db 473 TGGGCGGCTCGGCGATCTGACCCGCTTCGCGCATATGCTGCTGGCGCTGCGGAGCGG 532
Qy 655 GAGACTTCTTGGACCGGAGCGGACCGGCTTGAACGCGCGAGAGGCTCCGCGCGGAG 714
Db 533 GCGGATGATTCGACCCCTCGGCGCGGCTGCGAGGCGCGGCGGCTGATGATCGGCTCT 592
Qy 715 GCGTCAACCGCTCGATCTCTCCATCGGATGCACTTGGCGCTGCTCAACGCGGACCTTCG 774
Db 593 GCGGCGGCTCGCTGACGCTGCGCGCGCGGCTGACGCGGCTGGCGCTGGGATGCGACCTCG 652
Qy 775 CCATGACCGGAGTCGCGCTGGTGAATGCTCAGCGCTTCGCGCCATCTCGGCAACTCGGCGG 834

Db 653 CGATGACGGGATTCGGCCCTGACCGGGTCGAGCGCGCGGCGATCGACCGCGC 712
Qy 835 TGGCGTTGACCGCCCTGTTGCGGAATGTCGAGAGCGCGGACCGAGGCTATGGCGCGG 894
Db 713 TTCGGCACAGCGCGGCTCTGATGAGGTCTTGTCCGGTCTATGCCGAAAGCTTGGCATCGG 772
Qy 895 CACTGTCGACCTCGCGCGCATCCCGACAGAGAGCGCGGACGAGGCTGCGCGCC 954
Db 773 CTTTCCGAGAGCTCGCGCGCATCCGGGCGAGTGCAGGCGACCGAGCGGCTGCGGAGG 832
Qy 955 GCGTGGACGCGCAGCGCGGGTGTCTCGGCACAGCTATTGCGAGCGAGGCTCGACCGC 1014
Db 833 CGCTGGATGGCGCGGGGCTCTGTGCGACCTGACCGCGCGCGGCTGACCGCG 892
Qy 1015 GCGATATCGGACCGGAGCGCGGCGGAGAGTCTTACAGCCTGAGCGCTGCGCTCGC 1074
Db 893 CGGATCTCGCGCCCGAAGATCATCCGCGCAGAGTGTCTACAGTCTGCGGTGTGCGCG 952
Qy 1075 AGGTTCTCGGGCGGGCTTCGACACGCTTCGATGGCATGACCGGGTGTGACGATCGAGC 1134
Db 953 AACTGGTTCGGCGGGTCTGCGACACGCTGGAGCTGGACGATCGGTCTGCTCACCTGCGAGC 1012
Qy 1135 TGAACGCGGTGACGACAAATCCGGTGTTCGCGCCGATGGCAGCGTGC CGCGCCCTGCAACG 1194
Db 1013 TCAATTCGTCACGACAAATCCGATCTTCCGAGGGCTGCGCGGTGCGCGCCCTGCAACG 1072
Qy 1195 GGGGCAATTTATGGGCGAGCATGCGCTGAGCTGCGATGCGCTCGCCACCGCGCGTCA 1254
Db 1073 GCGGCAATTTATGGGCGGTGATGCTGCGCCCTTGGCTCGGATGGCTGAAACGCGCGCTGG 1132
Qy 1255 CCGTTCTGGCGGGCTTCGAGAGCGCGAGATTGACGCTGACAGATGAAGGCTGAAC 1314
Db 1133 TGACGCTGGCGGGCTGCTGAGCGGTGAGATCGCCCGGCTGACGACGAAAGCTGAACA 1192
Qy 1315 GTGGGCTGCCCCCTTCTCACCGGGCGCCCGCGGCTTGAATTCGGGCTTCATGGGG 1374
Db 1193 AGGCTCTGCGCGCTTCTGCAAGGGGGGAGGGCGGGCTGCAATCGGGCTTCATGGGG 1252
Qy 1375 CACAGGTGACGGGACCGCGCTCTGCGCGAGATGCGAGCCA---CGGGACCTGCTCTGA 1431
Db 1253 GCGAGGTGACGGGACCGCGCTTCTGCGGAAATGCGGGGAAATGCGCACCGCGTTCGG 1312
Qy 1432 TCCATTCGATCTCACGAAACCGCGCAATCGAGATGAGTGTGCTGCTGGGACGATCGCG 1491
Db 1313 TGCACTGCTGTTCACCAATGCGCGCAATCAGGATGTGCTGCTGATGGGAAAGATTGCG 1372
Qy 1492 CGGCGCTCTGCGGAGAGATCGACCGTTGGCGGAGATCCTTGGCATCTCGCTCTCT 1551
Db 1373 CGCGAGGGCGCGGCGCGAGTGTGCTGCCCTGTGCGAGATCCAGCGCATCTGCGGCTTG 1432
Qy 1552 GTCTTGACAAAGCTGCGGAGTGTGCTGCGGAGC---GGGCTAGACGGGCTGTCTCCCG 1608
Db 1433 CCCTTGCCAGCGATGATCTGCTGACGACCCCGAGGGGCGAGCGGATGCTGCTTA 1492
Qy 1609 CGGGGAAGAGCTGTGAGCGCCCTGCGAGAGATGTTCCCGCGCTTGAAGAGGACCGCG 1668
Db 1493 CGGCGCGGATCTGCGGAGACCGGATCGGGCGGTCTCGCGCGGGCTTCGCGCGCGACAG 1552
Qy 1669 CCCTGGAGAGGAATTCGCGGCTTGTGAGCACTTGTGAGCAATCTCCCGTC 1724
Db 1553 CGCTTCCGGGGATATCGAAGCGGTGGCACAGGCTGTGCGTCTCCCTCCNNNGCC 1608

RESULT 5

US-10-369-493-32499
; Sequence 32499, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 32499
; LENGTH: 1587
; TYPE: DNA
; ORGANISM: Ralstonia metallidurans
US-10-369-493-32499

Query Match 13.2%; Score 244.2; DB 6; Length 1587;
Best Local Similarity 50.2%; Pred. No. 3.9e-54;
Matches 780; Conservative 0; Mismatches 743; Indels 30; Gaps 6;

Qy 178 AGCGGCGCTCGAGTGGATCGGCACATCGATCTGGACAGGCCCATGCGGTGGGAGCG 237
Db 2 ATCGGCGGCATCGACGCGCCACACCTGACCCCGACACCGTCCGCGCCATCGCAGCG 61
Qy 238 GCGGCGCGGATTGTCTTCCCTCCGCGCGCGACCGGTGCGGTCCGTCGCGAAGCGC 297
Db 62 GCAGCGCGCGCATCGTCCGAGCCCGTCTCGGCAAGTGTGCGATGCCCGCGCC 121
Qy 298 GGCTCGGCGCTGTATCCGCGAGCGCGCATGTCTACGAGCTGACAAACCGGCTTCGTC 357
Db 122 GCTTCGAGCAGGTGGCTGCGGCCAAATGTGCGCATCTACGCGCTCTCACGGCTTTGGC 181
Qy 358 CCCTTGGGAACCGCTGATCTCAGGTGAGAAATGTCCGAACTGTCGACGCGCATCTTGTCC 417
Db 182 AACTGTGTACAACTGGGTGACATCGAACATGCGCGTGCCTGCGAGGAGAACCTGTTC 241
Qy 418 ATCATCTGGCCAGCGGCGTGGGACCGGTGTCTGACTGAGACGCGCGCGGCCATGTTTC 477
Db 242 GAGCCATTGCGGGGTGTGGGTCCGCTGTTTCGCGGAGGAGTCCGCGCGATGAG 301
Qy 478 TGGCGGCTGTGTGTCGATCGCTCAGGAGCGCTCCGTTGCGAGGAGGACCATCGCTC 537
Db 302 TCGCGGTGCAATGCTTGGCACGCGGATCTCGCGGTGCGCGCGCTTATCGAAC 361
Qy 538 GCCTGATCGACCTGCTCAATTCGAGTCTGCTCGCGCGCTTCCAGCGCGCGGACGTTGG 597
Db 362 AACTGTGAAATGATCTGGAAGCGGACATCGCGAGCGCTGCGCGAGTGGGTTTCGTCG 421
Qy 598 GCGCGTGGGTGACCTGACACCGCTTGGCATATGTTGCTCTGCTCCAGCGCGGGAG 657
Db 422 GGGCCAGCGGTGATCTCGCGCTCTGTGCGCATGTCGCTCATCGCTGATCGCGAGGCA 481
Qy 658 ACTTCTTGACCGGGACCGGACCGGCTTGACCGCGCAGAAAGGCTCCGCGCGGACGCG 717
Db 482 AGGTGCTGACCGAGGATGGCGGTACGCCACCCACGCGCGAAGTCTCTGCGGAGGTGGCA 541
Qy 718 TGCAACCGCTGATCTCTCCATCGCGATGCACTGGCGCTGTGTCACAGGACCTCCGCA 777
Db 542 TCACGCGCTGCGGCTGCGCATACAAAGGAGGGCTGGCGCTGATCAACCGGACATCGGCA 601
Qy 778 TGACCGGATCGCGTGGTGAATGCTCACCGCTGCGCCATCTCGGCACTGGGCGGTGG 837
Db 602 TGACCGGGGTGTGTGCTGTGTGGAGACGTGTCGCGCGCAGGTCCAGCAGGCCGAGA 661
Qy 838 CGTTGACCGCTGTGTGCGGAATGTCTGAGAGCGCGGACCGAGGCAAT---GGCGCGCG 894
Db 662 TCATCGGGGCTTGGCGCTCGAAGGATATCCGCTCGCGCATGCTTTCATGGCCCATG 721
Qy 895 CACTGTCCGACCTGCGCGCGCATCCCGGACAGAAAGAGCGCGCAGCAGGCTGCGCGCC 954
Db 722 GGCACGACATCGCCAAACCGCATCCGGAACAGATCCGCTCGGCGCGGCAACATGCGCGCG 781
Qy 955 GCCTGACGCGAGCGCGCGGTGTCGGGACG-----TCATTGCGCGGAGGCTCG 1008

```
Db 782 TGCTGCGCGATTGCGGACAGCGCTCTCCGGACATGGCGAACTGTCCGCGAGATGAAGACAC 841
Qy 1009 ACGCCGCGCATATCGGACAGCGAGCCGAGCGGG-----CAGGATGCCCTACAGCCTGC 1062
Db 842 GCGGGGCGAGGCAAGAACACCGGCACTGGCGTTCATCCAGAGGCTTACACGCTGC 901
Qy 1063 GCTGCGCTCCGAGGTTCTCGGGCGGGCTTCACACAGCTCGCATGACCGGGTGC 1122
Db 902 GCTGCATTCGCGAGGTGCTTGGCGCGGTGCGGATACGCTCGACCATGTCGCCACCGTGG 961
Qy 1123 TGACGATCGAGCTGAACCGGTGACGACAATCGGTGTTTCCGCCGATGGCAGGTGC 1182
Db 962 TCGAGCGCGAGTAATTCATCAATGACAATCGCTGTTCTTCGAAGACGGCGAGCTGT 1021
Qy 1183 CCGCCCTGACCGGGGGCAATTCATGGCCAGCATGTGGCGCTGAGTCCGATGCGCTCG 1242
Db 1022 TC-----CAGGGGGCACTTCACGGCCAGCAGGTGGATTTCGCAATGGATTCTCTGG 1075
Qy 1243 CCACGGCGGTCAACGTTCTGGCGGGCTTTCGGAGCGCCAGATTGCAAGTCTGACAGATG 1302
Db 1076 CCATCGCGGCAACGCAACTGGCGGTGGTTCGGAGCGCGCTGAACCGCTCTGCTGAGCC 1135
Qy 1303 AAAGGCTGAACCGTGGGTGCCCCCTTCTTCACCGGGCCCCCGGGTTGAATTCGG 1362
Db 1136 CGCATCTGAACAACAATCTGCGGCGCTTCTGCGGCGCGGAACGAGGGGTTGCTGTCG 1195
Qy 1363 GCTTTCATGGGCGCACAGGTGACCGCGACCGCTCTGCGCGGAGATGCGAGC---CACGG 1419
Db 1196 GGTTCGCGGGGACAGTATCCGGCCACGGCGTGTGATTCGGAGAACCGCACGATTCGCA 1255
Qy 1420 GACCTGCTCGATCCATTCGATTCACAGAACCGCCCAATCAGGATGTGCTCGCTTG 1479
Db 1256 GCCCGCGAGCATCCAGAGTGTCCGTCGAACCGCGACAAACAGGATGTGTCAGCATGG 1315
Qy 1480 GGACCATCGCGCGCGCTTTCGCGGAGAGATGACCGTTGCGGGCGAGATCCTTCGGA 1539
Db 1316 GGCTGATCGCTGCCCGCAATTCGCCCGCGCATTCGACAAATACCAAGTACATCCTCGCG 1375
Qy 1540 TCCTCGCTCTCTGCTTTCGACAAAGTTCGAGATGCGCTGCGCGACGGCTAGACGGG 1599
Db 1376 TGGAGTTGTTGGCGTCATGTACGCGCGCGAACT-----CGCGGGCGGTCGAGCAAC 1429
Qy 1600 TGTCTCCCGCGGGAAGAAGCTGGTCAGGCCCTGCGCAGCAGATTCCCGCGCTTGAGA 1659
Db 1430 TGGCGCGCGCAGCGCGCGTGTTCGCTGTCGCGGAGCGCGTCCGTTCTCTGTCGA 1489
Qy 1660 CGGACCGGCCCCGAGGAGAAATTGCGCGCTTGTCTAGGACCTCTTCGAG 1712
Db 1490 TCGATCGCTATATGACCGAGCATCGAGGCTATGCGCGGCTGCTCCGCTCAG 1542
```

RESULT 6

```
US-10-159-257A-1/c
; Sequence 1, Application US/10159257A
; Publication No. US20040161828A1
; GENERAL INFORMATION:
; APPLICANT: LIU, WEN
; APPLICANT: LIU, WEN
; APPLICANT: CHRISTENSON, STEVEN D.
; APPLICANT: STANDAGE, SCOTT
; TITLE OF INVENTION: GENE CLUSTER FOR PRODUCTION OF THE ENEDIYNE ANTITUMOR
; FILE REFERENCE: 407T-896020US
; CURRENT APPLICATION NUMBER: US/10/159,257A
; CURRENT FILING DATE: 2002-05-31
; PRIOR APPLICATION NUMBER: 09/478,188
; PRIOR FILING DATE: 2000-01-05
; PRIOR APPLICATION NUMBER: 60/115,434
; PRIOR FILING DATE: 1999-01-06
; NUMBER OF SEQ ID NOS: 207
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 42000
```

```
; TYPE: DNA
; ORGANISM: Streptomyces globisporus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (5982)..(7475)
; OTHER INFORMATION: orf(-2)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (13012)..(14076)
; OTHER INFORMATION: orf10
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (15901)..(15960)
; OTHER INFORMATION: splice variant a
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (25815)..(27167)
; OTHER INFORMATION: orf20
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (27214)..(28590)
; OTHER INFORMATION: orf19
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (29632)..(31194)
; OTHER INFORMATION: orf2
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (31280)..(32587)
; OTHER INFORMATION: orf3
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (32809)..(34389)
; OTHER INFORMATION: orf4
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (36165)..(37487)
; OTHER INFORMATION: orf5
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (37559)..(38935)
; OTHER INFORMATION: orf23
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (38983)..(39261)
; OTHER INFORMATION: orf26
; US-10-159-257A-1
```

```
Query Match 13.1%; Score 242.4; DB 7; Length 42000;
Best Local Similarity 50.3%; Pred. No. 1.1e-53;
Matches 749; Conservative 0; Mismatches 706; Indels 33; Gaps 5;

Qy 227 CGTGGGAGCGGCGCGCGGATTTCTTGGCCCTCCGGCGCGGACCGGTGCGGTGC 286
Db 40906 CGTCGCGGAGGAACGCGCGACGCTGACGTACCGGCGGAATCCATCGGAAGCCCGAGAA 40847
Qy 287 GTCCGAAGCGCGCTCGGCGCTGTCATCCGCGAGGCGGCCCATGCTACGACTGACAAC 346
Db 40846 GAGCCGGGAGATCTTCGAAGGGATGCCGAAACAGAAATCCCATCTACGGGTGACCAC 40787
Qy 347 CGGCTTCGCTCCCTTCGGAACCGCTGATCTCAGGTGAGAATGTCCGAACGCTGCAGGC 406
Db 40786 CGGGTACGCGAGATGATCTACATGCAAGTTCGACAAAGTTCGAAGGAGTTCGACAGAC 40727
Qy 407 CAATCTTGTCCATCATCTGGCCAGCGGCGTGGGACCGGTGCTTGTGACTGACGACGCGCG 466
Db 40726 CAATCTCGTCCGTAGCACACAGCGCGGGAGTCCGCTGCTTTCGCCGAGGACGAGCGCG 40667
Qy 467 CGCCATGTTCTTGGCGCTGCTGTCGATCGCTCAGGAGAGCTCCGCTGCCAGCGAGGG 526
Db 40666 GCGGATCGTCCGCGCGCGGCTGAAACACCTCGCAGAGGCCACTCCGCGGTGCGCCCAT 40607
Qy 527 GACCATCGCTCGCTGATCGACCTGCTCAATTCCGAGCTCGCTCCGGCGGCTTCCACGCG 586
```

Db 40606 CATCTCGAACCTTCGCGAGTACCTGAACAGGGCATCACCCGGCCATACCCGAGAT 40547
Qy 587 CGGACCGTGGCGCGTGGGTGACCTGACACACCGGTTGGCGATATGTTGCTTCGCTCCA 646
Db 40546 CGGGTCACTCGGGCGAGCGCGACCTGGCTCCCTCTCCACGTCGAGCACCTCAT 40487
Qy 647 GGGCCGGGAGACTTCCTGGACCGGAGCGGACGCGGCTTGACCGGCAGNAGGCTCCG 706
Db 40486 CGGAGGGCTACCTGCTCTG---CGGACGGACGGCGGTGGAGACCGCCAGGTGTGGC 40430
Qy 707 GCGCGAGCGGTGCAACCGCTCGATCTCTCCCATCGGATGCACTGGCGCTGGTCAACGG 766
Db 40429 CGAGCGGGCATCGACCGCTCGAAGTGGCTTCAAGAGGGCTCGCACTGATCAAGG 40370
Qy 767 GACCTCGGCATGACCGGGATCGCGCTGGTGAATGCTCACGCTTGGCCCATCTCGCAA 826
Db 40369 CACGTCGGGATGACCGGTCTGGGCTCCCTGGTGGAGCGGCCCTGGAGCAGGCCA 40310
Qy 827 CTGGCGGTGGCTTGAGCGCCCTGCTTGGGATGCTGAGAGCGGACCGAGGCGATG 886
Db 40309 GCAAGCGAGATCGTAGCGCTGCTCATGAGCGGTACGCGGATCGACCCAGCCCTT 40250
Qy 887 GCGCGCGCACTGTCGACCT---GCGCGCATCCCGACAGAGAGCGCGCAGCGAG 943
Db 40249 CCTCGGAGGGGACACATAGACCCCGCCGACAGCGCCAGATCGACACCGCGCAA 40190
Qy 944 GCTGCGCGCGCGTGGACGCGCGGCGGTGCTCGGCACGTCATGCGCGAGCGGAG 1003
Db 40189 CATGCGGCGCTGATGCGGGGACGCGACTGACGCTGAGCAGCAGCGCTGCGCGAGA 40130
Qy 1004 GCTC-----GACGCGCGGATATCGGACGAGCGCGAGCGGCGGCGAGGA 1048
Db 40129 ACTCCAGAGGACAAAGGAGCGCGCAAGGAGCTCAGCGCTCGGAGATCTACCTGCAGAA 40070
Qy 1049 TGCTACAGCTGCGCTCGCGCTCGCAGGTTCTCGGGCGGCTTCGACAGCTCGCATG 1108
Db 40069 GGCCTACTGCTGGGCGCATCCCGGAGTCTCGGGCGGCTGCGCGACACCTTGTACCA 40010
Qy 1109 GCATGACCGGCTGTGACGATCGAGCTGAACCGGCTGACGACAAATCCGCTGTTCCGCC 1168
Db 40009 CGCGCGGCACAGCTGCGCATCGAGCTCACTCGGCCAAAGCAACCGCTCTT----- 39956
Qy 1169 CGATGCGAGCGTGCCTCGCTGACGCGGGGCAATTTTCATGGCGCAGATGTCGCGTGAC 1228
Db 39955 CTTGAGGGCAAGGAGATCTTCCACGGGCGAACTTCCACGCTCAGCGGATCGCGTTCCG 39896
Qy 1229 GTCCGATGCGCTGCGCAGGCGCTCACCGTTCTGGCGGGCTTTCGAGGCGCCAGATGC 1288
Db 39895 GATGGACTTCGTGACCATCGCGCTCACCCAGCTCGCGCTCTTGGCCGAGCGGAGATCAA 39836
Qy 1289 ACGTCTGACAGATGAAGGCTGAACCGTGGCTGCGCCCTTCTTCCACCGGGCGCCCGC 1348
Db 39835 CGGGTCTGAAACCGGCACTTACGTACGCGCTCCCGAGTTCTCTGTCGCGGGACCC 39776
Qy 1349 CGGTTGAAATTCGGCTTTCATGGCGCAGAGTGACGCGCAGCGCTCTCTGCGCGAGAT 1408
Db 39775 GGGCTGCAAGCGGATTCGCGCGCGCCAGTACCGCGCCACCGCACTGGTGGCGGAGAA 39716
Qy 1409 GCGAGCCACGGGACCTGCTCGATTCATTCGATTCACGAAACCGCGCCATTCAGATGT 1468
Db 39715 CCGGACGATCGCGCGCGCCAGCACCCAGAGGCTCCGCTCCAAACCGCGCACACAGGACGT 39656
Qy 1469 GGTCTGCTTGGGACCATCGCGCGCGCTTTCGCGGAGAGATCGACCGTTGGCGGA 1528
Db 39655 GGTGAGCATGGGCTGTATCTCGGCGCGCCGAAACCGCCCGGCTCTGTGCAACAAACAA 39596
Qy 1529 GATCCTTGGCATCTCTGCTCTGTCACAGCTGCGGAGTGGCGCTGCGGCGAGCGG 1588
Db 39595 GATCCTCGCG-----GTGGAGTACCTGGCGCGCGCCAGGCGGTTCACATCTCCGCGCG 39542
Qy 1589 CCTAGACGGGCTGTCTCCCGCGGGGAAGAGCTGGTGCAGCGCCCTGCGGAGCAGTTCCC 1648

Db 39541 GTTCGACGGCTTGAGCCCGGGCGAAGCCACGCTACGAAGCGGTGCGCGCTGTTCC 39482
Qy 1649 GCCTCTTGAGACGAGACCGGCCCTTGGGACAGAGAAATTCGCCCGCTTGC 1696
Db 39481 GACGCTGGGCTGACACCGGTACATGTCGCGGACGACATCAGCTGGTGC 39434
RESULT 7
US-10-292-198-1/c
; Sequence 1, Application US/10292198
; Publication No. US20030157654A1
; GENERAL INFORMATION:
; APPLICANT: SHEN, Ben
; APPLICANT: LIU, Wen
; TITLE OF INVENTION: BIOSYNTHESIS OF ENEDIYNE COMPOUNDS BY MANIPULATION OF C-1027 GENE
; FILE OF INVENTION: PATHWAY
; FILE REFERENCE: 054030-0007
; CURRENT APPLICATION NUMBER: US/10/292,198
; CURRENT FILING DATE: 2003-03-14
; PRIOR APPLICATION NUMBER: US 10/159,257
; PRIOR FILING DATE: 2002-05-31
; PRIOR APPLICATION NUMBER: US 09/478,188
; PRIOR FILING DATE: 2000-01-05
; PRIOR APPLICATION NUMBER: US 60/115,434
; PRIOR FILING DATE: 1999-01-06
; NUMBER OF SEQ ID NOS: 146
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 63158
; TYPE: DNA
; ORGANISM: Streptomyces globisporus
; US-10-292-198-1

Query Match 13.1%; Score 242.4; DB 6; Length 63158;
Best Local Similarity 50.3%; Pred. No. 1.1e-53;
Matches 749; Conservative 0; Mismatches 706; Indels 33; Gaps 5;
Qy 227 CGTGGCGAGCGCGCGCGGATTTGCTTGGCCCTTCGCGCGCGCGACCGGTGCGGTGC 286
Db 40901 CGTGGCGAGAAACGCGCGACGCGTTCGACGTACCGCGCAATCCATCGGAAGGCCAGAA 40842
Qy 287 GTCCGAAGCGCGCTCGCGCTGTCATCCGCGAGGCGCGCATGCTTACGCACTGACAAAC 346
Db 40841 GAGCGCGGAGATCTTCGAAGGGATTCGCGCAACAGAAATCCCCATCTACGCGGTGACCA 40782
Qy 347 CGGCTTCGCTCCCTTCGGAACCGCCCTGATCTCAGGTGAGAACTCGGAACGCTGCAAGC 406
Db 40781 CGGCTACCGGAGATGATCTACATGAGGTGCAACAGTCGAAGAAAGTCGACTGACAGAC 40722
Qy 407 CAATCTTTGTCATCATCTGGCCAGCGCGTGGGACCGGTGCTTGACTGGACGACGCGCG 466
Db 40721 CAATCTCGTCCGTAGCCACAGCGCGGAGTCGGTCCGCTGTTCCCGAGGACGAGGCGCG 40662
Qy 467 GCGCATGGTTCGCGCGCTCTGGTGTGATGCTCAGGAGGACCTCCCGGTGCCAGCGAGGG 526
Db 40661 GCGCATGCTGCGCGCGCGCTGAAACCCCTCGCAAGGCGCACTCCCGGTGCGCCCAT 40602
Qy 527 GACCATCGCTCGCTGATCGACCTGCTCAATTCGAGCTCGCTCCGCGCGCTTCCAGCGCG 586
Db 40601 CATCTCTGAACGCTTCGCGCAGTACCTGAAACGAGGGCATCACCCCGGCCATACCCGAGAT 40542
Qy 587 CGGCAAGGTGGGCGCGTTCGGGTGACCTGACACCGCTTTCGCGCATATGGTGTCTTCGCTCCA 646
Db 40541 CGGCTCACTCGGGCGGAGCGCGACCTGGCTCCCTCTCCACGTCGCGAGCACCCCTCAT 40482
Qy 647 GGGCGGGGAGACTTCTTGAGACCGGGACGCGGACGCGGCTTGAACGCGCGAGNAGGCTCCG 706
Db 40481 CGGAGAGGGCTACGTCTCTG---CGCGACGAGCGCGCGGTGGAGACCGCCAGGTGCTGGC 40425
Qy 707 GCGCGGACGCTGCAACCGCTCGATCTCTCCATCGGATGCACTGGCGCTGGTCAACGG 766
Db 40424 CGAGCGGGGATCGAGCGCTCGAACTGCGCTTCGAAGGAGGCGCTCGCACTGATCAACGG 40365

Db 791 AGAGCCACCGTGTAGGCGGACC-----AGC 814
Qy 1039 CGGGGAGGATGCTACAGCTGTGGCTCGCGAGGTTCTCGGGGCGGGCTTCGACA 1098
Db 815 GGGTGCAGAGACCCCTACTGTATCCGCTGCCAGCGGAGGTGACGGGCGCGCATGTGATG 874
Qy 1099 CGCTCGCATGTCATGACCGGTGTGTGACGATCGAGCTGAAACGCGGTGACCGACAAATCCGG 1158
Db 875 TGCTGCGATGCGGCGGGACAGCTGGCCACCGAGGCCAATGCCCGCCACCGACAAATCCGC 934
Qy 1159 TGTTCGCGCCGATGTCAGCGGTGCGCCCTTCGACAGGGGGCAATTTTCATGGGCGCAGCATG 1218
Db 935 TCGTGTCTCGGA-----CGGGCGCATGTCTCGGAGGCAACTTTCATGCGGAGCCCG 988
Qy 1219 TGGCGCTGACGTTCGATCGCTGCCACGGCGGTACCGTTCTTGGCGGGGCTTTCGAGAGC 1278
Db 989 TGGGCTTTCGCGCGCGACATGATCGCTGTGGCGCTCTCCGATCGGGGCCATCGCGCAGC 1048
Qy 1279 GCCAGATTGACGCTCTGACAGATGAAGGCTGAACCGTGGGCTGCCCGCCCTTCTCTCCACC 1338
Db 1049 GCGCGGTGGCGCTGATGGTGATCCGACGCTCTCTTCGACCTTCCGCGCTTCTCTCACCC 1108
Qy 1339 GGGGCCCGCGCGGTGAATTCGCGCTTCATGGGCGCACAGGTGACGGCGACCGCGCTCC 1398
Db 1109 CCGAGC---CGGGCTGAATTCGCGGCTGATGTCGCGGAGTGACAGCGGCGCGCTCA 1165
Qy 1399 TGGCCGAGATGTCGAGCCACGCGACCTTCCCTGATFCCATTCGATCTCCACGAACCGCCCA 1458
Db 1166 TGAGCGAGAACAGACATGGCGCCGCCACCGTCCACCGACAGCACGCCACCTCCGCCCA 1225
Qy 1459 ATCAG----GATGGTCTCGTTGGGACCATCGCGCGCGCTCTGCGCGAGAGATCG 1515
Db 1226 ATCAGGAAGATCATGTAGCATGGCGGCCCATGGCGCGCGAGGCTCGGCGCGGTGCTG 1285
Qy 1516 ACCGTTGGGCGGAGATCTTGGATCTCGCTCTCTGCTCTGTCACAGCTGCGGAGCTGC 1575
Db 1286 AGAACCTCGCGGTGATCTCGGACCGAGGGATCTCGCGCGCGCAAGGGGTGAGTTCC 1345
Qy 1576 GCTCGCGAGCGGCTAGAGCGGGTGTCTCCCGGGGGAAGAGCTGGTGAGGCCCTTCG 1635
Db 1346 GCGCGCCCTCGCCACCTCCGCCCGCTCGGCGCGGTGCTGGCGCGGTGGCGCGGAGG 1405
Qy 1636 GCGAGAGTTCCCGCC 1651
Db 1406 TGGCGCGATCGCGGC 1421

RESULT 9

US-10-156-761-3312
; Sequence 3312, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 3312
; LENGTH: 1542
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis

Query Match 12.6%; Score 232.8; DB 6; Length 1542;
Best Local Similarity 50.0%; Pred. No. 3.8e-51;
Matches 730; Conservative 0; Mismatches 682; Indels 48; Gaps 4;
Qy 215 CAGAGCCCATGCGGTGCGAGCGGCGCGCGGATGTCTTCCCTCCGCGCGCGA 274
Db 48 CGAGTCTTCGCGGTGCGGCGCGCGCGCGATCGAGCTCTCCGGGAGGCGGTGAC 107
Qy 275 CCGGTGCGGTGCGGTGCGAAGCGGCGCTCGGGCTGTCTATCGCGAGGCGCGCATGTCTA 334
Db 108 CGGCTTCGCGCGGCGCGCGCGCATCGGTGAGACGCGCTGGCGGCAAGCGGAGCCCGTGT 167
Qy 335 CGGACTGACAAACCGGCTTCGCTCCCTTGGGAACCGCTGATCTCAGGTGAGAAATGTCCG 394
Db 168 CGGGGTGAGCACCGGCTTCGCGGCGCTTGGCGACCCGCGCATCATCAGCCAGGAGCTCGGGC 227
Qy 395 AACGCTGACGGCCAAATCTTGTTCATCATCTGGGCGAGCGGCGTGGGACCGGCTGTGACTG 454
Db 228 GCAGCTCCAGCGCAACATCTGTCTCGCTCGCACGCGCGCATGGTCCGAGGGTGGAGCG 287
Qy 455 GACGACGGCGCGCGCCATGTCTTGGGCGCTGTGGTGTCTGATCGCTCAGGAGCTCCGG 514
Db 288 GGAGGTGCTACGGGCGCTGATGTCTTTCGAGCTGAAACCCGCTCTGCTCGGCGCACACCGG 347
Qy 515 TGCAGCGAGGGGAGCCATCGCTCGCTGTATCGACCTGCTCAATTCGAGCTCGCTCCGGC 574
Db 348 GTGACGCGCCGAAAGTTCGCGCAGAGATGCGCGCATCTCTCAACCGCGGATCACCCGCT 407
Qy 575 CGTTTCCGAGCGCGGACCGGTGGGCGCTCGGGTACCTGACACCGCTTGGCGCATATGCT 634
Db 408 CGTGACAGTACGGCTCGCTCGGCTGCTCGGCGACCTCGCCCGCTCTCCCACTGCGC 467
Qy 635 GCTTGTCTCCAGGCGCGGGAGACTTCTTGAGACCGGAGCGGAGCGGCTTGAAGGGCGC 694
Db 468 CCTGACGCTGATGGGCGAGGCGCAGCGGAGGGCCCGGACCGCACCGCTCGCGCGCGG 527
Qy 695 AGAAGGCTCCGGCGCGGAGCGGTGCAACCGCTGCTGATCTCTCCCATCGCATGCTGCG 754
Db 528 TGACCTCTCGCGAGCAGCGGATTCAGCGCGGTGAGCTGCGCGAGAGGAGGCGCTTGGC 587
Qy 755 GCTGCTCAACGCGGACCTTCGCCCATGACCGGGATCGCGCTGTGAAATGCTCACGCTGCGC 814
Db 588 CCTCTCAACGCGCACCGACGCGCATGCTCGCATGCTGCTCATGCGCCCTCGCGCACTGGA 647
Qy 815 CCATCTCGGCAACTGGGCGGTGGGTTGACGGGCTCTGCGGCAATGTCTGAGAGCGCG 874
Db 648 CGGCTCTTACAGTCTGCGGACATCACGGCGGCGCTCTCCCTCGAGGCACTCTCTCGGCAC 707
Qy 875 GACCGAGGATGGGCGCGGCACTGTCCGACCTGCGGCGCATCCCGGACAGAGGAGCGC 934
Db 708 CGACAAAGTGTCTCGCCCCCGAATTCAGCACCATCCGCGCGCACCCCGCTCAGGGCGCTC 767
Qy 935 CGCAGGAGGCTCGGCGCGCGCGTGGACGCGAGCGCGCGGTGTGTCGCGCACGCTATTGC 994
Db 768 GCGCGCCAACTGTGCGCGTACTGCGGGGTTCGAGC-----805
Qy 995 CGAGCGGAGGCTCGAGCGCGCGGATATCGGGAACGAGCGGAGCGGCGGCGAGGATGCTTA 1054
Db 806 -----TCACCGGCCACCAACAGGACGACGCGCGCGCTCCAGGACGCGGTA 851
Qy 1055 CAGCCTGCGCTCGCTCCGCAAGTTCTCGGGGCGGGTTCGACACGCTCGCATGCGCATGA 1114
Db 852 CTCCTGCTGCTGCGCGCGCGAGTTCGCGCGCGCGCGCGGACACCTTCGCGCCACGCGCG 911
Qy 1115 CCGGGTGTGACGATCGAGCTGAACCGGCTGACCGCAATCGGTGTTTCCGCGCGGATGG 1174
Db 912 CCTCTGCGCGAGGCGAGCTGGGCTCGGCGGTGGACAAACCGGTGTGTCTCCCGA---968

QY 1175 CAGCGTCCCGCCCTGCACGGGGGCAATTTTCATGGGCCAGCATGTGGCGTGCAGTCCGA 1234
Db 969 ---CGGGCGCGTGCAGTCCAAACGGCAACTTCCACGGTGCCTCGCGTACGTCCTCGA 1025
QY 1235 TGGCGTCCGCACGGCGCTCACCGTTCTGGGGGCGCTTGGGAGCGCCAGATTGCAGTCT 1294
Db 1026 CTTCTCTCGGATCGCGCGGCCACTCGGGTCCATCGCGAGCGCCGACCGCGGT 1085
QY 1295 GACAGTGAAGGCTGAACCGTGGGCTGCCCGCTTCTCCACCGGGGGCCCGCGGGTT 1354
Db 1086 TCTCGACAAGACCGCTCGCACGGCTGCGCGCTTCTCT---CGCGAGACGCGGCGT 1142
QY 1355 GAATTCGGGCTTATGGGCGCACAGGTGACGGCGACCGCGCTCTCTGGCGAGAT---GGG 1411
Db 1143 CGACTCCGGCTGATGATCGCCCACTACACGAGCGGGCCCTGGTCAGCGAGATGAAGCG 1202
QY 1412 AGCCACGGGACCTGCTCGATCCATTTCGATCTCCAGAACCGCCCAATCAGGATGTGT 1471
Db 1203 GCTCGGGTCCCGCGCTCGCGACTCGATCCCGTCTCGGCGATCGAGAGGACCACT 1262
QY 1472 CTCGCTTGGGACCATCGCGCGCGCTCTGCGCGAGAAAGATCGACCTTTGGGCGGAGAT 1531
Db 1263 CTCGATGGGTGGTTCGCGCGCGCAAGCTCCGTACGGCGATCGACAACCTCAGCGCAT 1322
QY 1532 CTTTGGATCTCGCTCTCTGTCTTGACAAAGCTGCGGAGCTGCGTGCGBGAGCGGCT 1591
Db 1323 CGTGGCGATCGAGCTACTACGCCGCCACCGCGGCATCGAACTCCGCGAGGGTCTGACCCC 1382
QY 1592 AGACGGGGTCTCCCGCGGGAGAGCTGTGTGAGGGCCCTGCGGAGCAGTTCCCGCC 1651
Db 1383 CGCCCCGGGTGCAGCGGCTATCAGGCGCTCCGAAAGGGGCGTGAAGGCCCAAG 1442
QY 1652 GCTTGAGACGGACCGGCCCC 1671
Db 1443 ACGGACCGCTTCTGGCCC 1462

RESULT 10

US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1
Query Match 12.6%; Score 232.8; DB 6; Length 9025608;
Best Local Similarity 50.0%; Pred. No. 3.6e-51;
Matches 730; Conservative 0; Mismatches 582; Indels 48; Gaps 4;
QY 215 CAGAGCCCATGCGTGGCGGCGGCGGAGTGTCTTGCCCTTCCGCGCGCGCA 274

Db 4139678 CGACGTTCTCGCGTGGCGCGTGGCGGCCCGGATCGAGCTCTCCGGGAGCGGTGAC 4139619
QY 275 CCGGTCCCGTGGCTTCGAAGCGCGGTTCGGCGCTGTCATCCGCGAGGCGCGCATGTCTA 334
Db 4139618 CGCCCTCGCGGCGCGCGGATCGTGAACGCGCTGGCGGCCCAAGCCGAGGCCGTGTA 4139559
QY 335 CGGACTGACAACCGGTTCCGTCCTTCGAAACCGCTGATCTCAGGTGAAGATTCGG 394
Db 4139558 CGGGGTGAGCACCGGCTTCGGGGCCCTGGCGACCCCGGCACATCAGCCAGAGCTCGGGC 4139499
QY 395 AAGCTGAGAGCAATCTTGTGCATCATCTGGCCAGCGGCGTGGGACCGGTGTTGACTG 454
Db 4139498 GCAGTCTCCAGCGCAACATCGTCCGCTCGACGCGCGCGCATGGGTCCGAGGGTGAGCG 4139439
QY 455 GAGCAGCGCGCGCATGTTCTGGCGGTCTGGTGTGATCGATCGCTCAGGAGCCCTCCGG 514
Db 4139438 GAGGTGCTACGGCGCTGATGTTCTGCGACTGAAAACCGTCTGTCGGGCCACACCGG 4139379
QY 515 TGCCAGCGAGGGGACCATCGCTCGCTCGATCGACCTGCTCAATTCGAGTCTCGTCCGGC 574
Db 4139378 CGTAGCGCCGAAGTCCGCGACAGCATGCGCGACATCTCAACGCGGGATCACCCCGT 4139319
QY 575 CGTTCCAGCCCGGACCGGTGGGCGCTGGGTGACCTGACACCGCTTTCGCGATATGTT 634
Db 4139318 CGTGACGAGTACGGTCTCGTCCGCTGCTCCGCGACCTCGCCCGCTCTCCACCTGCGC 4139259
QY 635 GCTTGCCTTCAGGGCGGGGAGACTTCTTGACCGGGACGGAGCGCGCTTGACGGCGC 694
Db 4139258 CTTGACGCTGATGGGAGGGCGACCGGAGGGCCCGGACGCCGCTCGTCCGCGCCG 4139199
QY 695 AGAAGGGCTCCGCGCGGACGGCTGCAACCGCTCGATCTCTCCCATCGGATGTCATGGC 754
Db 4139198 TGACCTCTCTCGCGAGCAGGCATCACGCGGTTCGAGCTGCGGAGAGGAGCGCTGGC 4139139
QY 755 GCTGTCAAAGGACCTTCGCCCATGACCGGATGCGGCTGTGTGAATGCTACGCTTCGCG 814
Db 4139138 CCTCTCAAAGCAGCAGCGCATGCTCGGCGATGCTGCTGCTGCTGCGCGCTCGCG 4139079
QY 815 CCATCTCGGCAACTGGGCGGTGGCTTGACGCGCTCTGTTGGGATGTCGTGAGAGCGCG 874
Db 4139078 CGCGCTTCAAGTCTGCGGACATCACGCGCGGCTCTCTCCCTCGAGGCACTCTCTGG 4139019
QY 875 GACCGAGCATGGCGCGGCACTGTTCGACCTGCGCGCGCATCCCGGACAGAGGACGC 934
Db 4139018 CGACAAGGTGCTCGCCCCCGAACTGCACGCCATCCGCCGACCCCGGTTCAGGGCG 4138959
QY 935 CGCAGGAGGCTGCGGCGCCCGGTGGAACGCGAGCGCGGGTGTTCGCGCACGTCATTGC 994
Db 4138958 GGCGCGCAACATGCTGGCGGTACTGCGGGGTTCCGAGC----- 4138921
QY 995 CGAGCGGAGCTCGACGCGCGCGATATCGGAGCGGAGCGGAGCGGCGGATGCGCTA 1054
Db 4138920 -----TCACCGGCCACACACGAGGACGCGCGCGGTTCAGAGCCGTA 4138875
QY 1055 CAGCTTCGCGTCCGTCAGGTTCTTCGGGGCGGGCTTCGACACGCTCGCATGGA 1114
Db 4138874 CTTCCGTGCGTCCGCGCGCGAGTCCGCGCGCGCGGACACCTTCGCCACGCCCG 4138815
QY 1115 CCGGGTGTGACGATCGAGCTGAACCGGTTGACCGCAATTCGGTGTTCGCGCCGATGG 1174
Db 4138814 CCTCGTCCGCGAGCGAGCTGGCGTTCGGCGGTGGAACACCGGTGTCTCCCGA--- 4138758
QY 1175 CAGCGTGGCGCGCTGCACGGGGCAATTTTCATGGGCCAGCATGTGGCGCTGACGTC 1234
Db 4138757 ---CGGCGCGTTCGAGTCCACCGCAATTTTCAGGTGCGCGCTGCTGCTCTCGA 4138701
QY 1235 TCGCGTTCGACGCGCGTCAACCGTTCTGGCGGCGCTTCGAGGCGCGCAGATTGCACTCT 1294
Db 4138700 CTTCTCGGATCGCGCGGCGCGACCTCGGGTCCATCGCGGAGCGCGCACCGCGGT 4138641
QY 1295 GACAGATGAAGCTGAACCGGTGGGCTGCCCGCTCTCTCCACCGGGGCGCGCGGTT 1354

Db 965 TGTGGGCGGCAATTTCCAGCGGAGCCGCTCGCTTCGGCGCGCAAAATCTCGCATCG 1024
Qy 1249 CGGTACCGTTCTGGCGGGCTTTCGAGCGCCAGATTTCAGCTCTGACAGATGAAGGC 1308
Db 1025 CCGCGCGCAGATTCGGCGGCTCGCGAGCGCGCATCGCGCTGTGTGATCGACGACGC 1084
Qy 1309 TGAACCGTGGGCTGCGCCCTTCTCCACCGGGGCCCGCGGGTTGAATTCGGCTTCA 1368
Db 1085 TCTCCGCGCTGCGCC-----TTTCTCGTGAAGNAGCGCGCTGAACTCGGGTTCA 1138
Qy 1369 TGGGCGCAGGTGACGGGACCGCGCTCTGCGCCGAGATTCGAGC---CAGGGACCTG 1425
Db 1139 TGATCGCAGCAGTGAAGCGCGCGCTCGCGTCGGAAACAGACGCTCGCGCATCCCG 1198
Qy 1426 CCTCGATCATTCGATCTCCACCAAGCCGCCAATCAGATGTGCTCGTTGGGACCA 1485
Db 1199 CGTCGTCGATTCGCTGCGCAGCTCGCGAATCGCGAAGCAACAGTGTGATGGCGACT 1258
Qy 1486 TGGCCGCGCGCTCTGCGCGAGAGATCGACCGTTTGGCGGAGATCTTTCGATCCTCG 1545
Db 1259 TCGCGCGCGCAAGCTACCGGACATCGGAGAACTCGGAAACATCTCGCATCGAGC 1318
Qy 1546 CTCCTGTCTTGACAAAGCTGCGGAGCTGCGCTGCGGCGAGCGGCTTAGACGGGTTCTC 1605
Db 1319 TGCTCGCGCGCGCAAGCGCTCGACTGC-----GCGCGCGCAACGCAACGAGCC 1369
Qy 1606 CCGCGGGGAGAGCTGTGCGAGCCCTCGCGAGCAGTTCCGCGCTTGAGACGGACC 1665
Db 1370 CGGCGCTGACGACCGGATGAAGCAATTCGCGCGGACGTCGCGGCACTACGATCTCGACC 1429
Qy 1666 GGCCCTGGGACAGGAAATTCGCGCGCTTGCTACGCACCT 1705
Db 1430 ACTACTTCGCGCGCGACATCGCGTGGTTCGCGCGCGCT 1469

RESULT 12

US-10-369-493-42898
; Sequence 42898, Application US/10369493
; Publication No. US2003023675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 42898
; LENGTH: 1503
; TYPE: DNA
; ORGANISM: *Mycococcus xanthus*
US-10-369-493-42898

Query Match 11.6%; Score 215.4; DB 6; Length 1503;
Best Local Similarity 50.0%; Pred. No. 1.4e-46;
Matches 753; Conservative 0; Mismatches 691; Indels 63; Gaps 6;
Qy 203 CATCATCTGGACCAAGCCCATCCCTGGGAGCGCGCGCGGCGGATGTCTTCCCCC 262
Db 21 CCTGAAGCTCGAGAAATCTCCAGTGGCTCGCAACGAGGCCACCGTGGAGCTGTCCG 80
Qy 263 TCGGCGCGCGACCGGTGCGCTCGTCCGAAGCGCGCTCGGGCTGTATCCGCGAGCG 322
Db 81 CGNAGCGCGCACCGGTGCGGCGCTCGCGCGCCCTGTGTGACCGGTGCGCGCGGAGA 140
Qy 323 GCGCCATGTCTACGGACTGACAAACCGGCTTCGGTCCCCCTTGGGAACCGCGCTGATCTCAGG 382

Db 141 CACGCCCGCTACTCGGCATCAACACCGGCTTTGGCAGCTTGGCGGAGGTCCGATCGACAA 200
Qy 383 TGAGATGTCCGNAACGCTCAGGCGCAATCTTGTCCATCATCTGGCCAGCGGCTGGGACC 442
Db 201 GAAGGACCTTGCAGCAGCTCCAGCGCAACCTCATCTCTCCACGCGTGTGGGCTCGGCAC 260
Qy 443 GGTGCTTGATCGACACCGCGCGCCCATGGTTCTTGGCGCGTCTGGTGTTCGATCGCTCA 502
Db 261 GCCCTCTCCCTTCCGGAAGCGGGGCGCTCTGTCTCGCTCCGCTGCAACGTGTCTGCCAA 320
Qy 503 GGGAGCTCCGCTGCGAGGAGGACCATCGCTCGCCCTGATGCACTGCTCAATTCGGA 562
Db 321 GGGTACTCCGCGATCCGATGAGACGCTGCGCTTGGCGCTGGAATGCTGAACCGGGA 380
Qy 563 GCTCGCTCCGCGCGTTCGCCAGCGCGGCGAGGTGGGCGGCTGGGTGACTGACACCGCT 622
Db 381 CGTGTGCGCGTGGTCCCCGAGCGGGGAGCGTGGGCGCTCCGGGATCTCCGCCCGCT 440
Qy 623 TCGCATATGTGCTCTGCTCCAGGGCCGGGAGACTTCTTGGAACGGGACGGGACGCG 682
Db 441 GGGGACCTTGGCGCTCGTCTTTCATCGGCGAGGTGAAGCTTCTATCAGGGCCA---GCG 497
Qy 683 GCTTGAAGCGCGCAGAAAGGCTCCGGCGCGGACGGCTGCAACCGCTCGATCTCTCCATCG 742
Db 498 GATGCCCGGAAAGCGGCGCTGGAGCGCGCGGCTTGCACACCGGTGGTGTGGAGGCCAA 557
Qy 743 CGATGCACTGGCGCTGCTCAACGGGACCTCCGCCATGACCGGGATCGCGTGGTGAATGC 802
Db 558 GGAGGCGCTCGCGCTTGTGTGAACGGGACACAGGCGCATGTGCGGCTGGGACACCTGTCCA 617
Qy 803 TCACGCTCGCGCCCATCTCGGCAACTGGCGGCTGGGTTGACGGCGCTGTTTCGCGAATG 862
Db 618 GCTTCGCGCGGAGTCCCTGCGGAGACATCGCCGAGCTCGCGGGCGCATGACGCTGAGGG 677
Qy 863 TCTGAGAGCGCGGACCGAGGCAATGGGCGCGGCACTGTCCGACCTTGGCGCGCATCCCCG 922
Db 678 GCTGTGGGAAGCCCAAGACCTTCAATCTGAGATTACGACGCTCGCGCGCACCCCGG 737
Qy 923 ACAGAGGACCGCGCAGCGAGGCTCGCGCGCGCTGGAGCGGCGCGCGGTGTGTCG 982
Db 738 CCAGAGGACGTCGCGCGCACCTCGCGCGCATCTGTGTGGAAGGAGCTGTGTGAGTC 797
Qy 983 GCACGTCAATTGCGGAGCGGAGGCTCGACCGCGCGCATATCGGACGAGCGGAGCGGG 1042
Db 798 GCAAGTCA-----ACTGCAAGAGT 818
Qy 1043 GCAGATGCTACAGCTCGCTCGCTCGCGAGGTTCTGGGGCGGGCTTCGACACGCT 1102
Db 819 GCAGGACCCCTACTCCCTCGCTGATGCGCGAGGTGCAAGCGCGCGCGCGAGGGCAT 878
Qy 1103 CGCATGGCATGACCGGCTGCTGACGATCGAGCTGAACCGGTGACGCAATCCGGTGT 1162
Db 879 CGGTTCTCCCGCGCATCTCGAGGTGAGGTCAACAGCGGACGGAACACCCGCTCGT 938
Qy 1163 TCCGCCCGATGGCAGCGTCCCGCTCGACCGGGGCAATTTCAATGGGCGAGCATGTGC 1222
Db 939 ---GTTGCGGACACGAGGCGCATCGTGTGCGGCGGCAACTTCCACGCGCAGCCATCTC 995
Qy 1223 GCTGAGCTCGATGCGCTCGCCACCGCGCTCACCGTTCTGGGGCGGCTTGGGAGGCCCA 1282
Db 996 CTTGGCATGAGCGTGTGGCGATGCGCGCTGACGCAACTGTCTCATCAGCGAGCGCG 1055
Qy 1283 GATTGCACTCTGACAGATGAAGGCTGAACCGTGGCTGCCCGCTTCTCCACCGGG 1342
Db 1056 CGTGAAGAGCTGTGGAACCCGCTGCTGTCAACCTGCGCGGT-----TCTCGCGAA 1109
Qy 1343 CCGCGCGGCTTGAATTCGGCTTTCATGGGCGCAGAGTGAAGCGGACCGCTCTCTGC 1402
Db 1110 GAATCCGGGTTGAATCTCGGCTTTCATGATCGCGCAGGTGACCGAGCGCGCTGTGGC 1169
Qy 1403 CGAGATGCGAGCCACGGGA---CCTGCGCTCGATCTCATTTGATCTCCACGACCGCCCAA 1459

Db 1170 CGAGTCCCGCGTGTGAGCCACCCGCGTCCGTGGATTGATTCCGTCATCCCGGGCGG 1229

Qy 1460 TCAGATGTGTGCTGCTTGGGACCATCGCGCGCGCTCTGCGCGAGAAATCGACCG 1519

Db 1230 AGAGCACCGTGTCCATGGGCATGACGGCGCGCTCAAGGGCGCTCAGGTTCAGCGACTT 1289

Qy 1520 TTGGGGGAGATCCTTGGGATCCTCGCTCTCTGCTCTGCTTTCGACAAAGCTGGGAGCTGGCTG 1579

Db 1290 CGCCGCTTGTGCTTGTGATGAAATCTCTGCTGGCGCGCAGGCCCTTGGACTTCCGCTT 1349

Qy 1580 CGGCAGCGGCTAGACCGGGTGTCTCCCGCGGGGAAAGCTGCTGAGGCCCTTGCAGCGA 1639

Db 1350 GCGCTGAAGCCCGCAAGGCGCTCGCGGCTACGAGCTGG-----TGGGCTC 1400

Qy 1640 GCAGTTCGCGCGCTTGAAGCGACCGGCCCTCGGACAGAAATTCGCGCTTGTGTAC 1699

Db 1401 GAAGGTCCCCACATGACAGGACCGAGCTGCACCGGACATCGAGGCGGTGAGCCA 1460

Qy 1700 GCACCTC 1706

Db 1461 GCTCGTC 1467

RESULT 13

US-10-369-493-28443

; Sequence 28443, Application US/10369493

; Publication No. US20030233675A1

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10369,493

; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039

; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 28443

; LENGTH: 1503

; TYPE: DNA

; ORGANISM: Burkholderia fungorum

US-10-369-493-28443

Query Match 11.4%; Score 211.6; DB 6; Length 1503;

Best Local Similarity 51.3%; Pred. No. 1.4e-45;

Matches 657; Conservative 0; Mismatches 569; Indels 54; Gaps 5;

Qy 301 TCGGCGCTGTATCCGCGAGCGCGCATGTCTACGACTGACAAACCGGCTTCGGTCCCG 360

Db 125 TCGCGGATATGCGCGAGGGCGAGCGGCTACGCGATCAACACGGGCTTTGGGCGCC 184

Qy 361 TTGCGAACCGCTGATCTCAGGTGAGAAATGTCCGAACGCTGCAGGCCAAATCTTGTCCATC 420

Db 185 TCGCCAGCAGCATATTCCGCAACACCACTCGAACTGTTGACGCGCAATCTGGTGTCT 244

Qy 421 ATCTGGCCAGGGGTGGGACCGGTGCTTGAATCTGGAACGCGCGCGCCCATGTTCTTG 480

Db 245 CTCACGGGTGCGGCTGGGTGAGCCGATGTCCGCTCGCGTCTGCTGTGATCGCGC 304

Qy 481 CGGCTGTGTGCTCATCGCTCAGGAGGCTCCGCTGCGGCGGAGGAGCCATCGCTCGCC 540

Db 305 TGAACCTCTGAGCTTCGCGCGCGCCCAATTCGGGCAATTCGCGGTAATGATGACGCGC 364

Qy 541 TGATCGACCTGCTCAATTCGAGCTCGCTCGGCGGCTTCCAGCGCGGCGACGCTGGGCG 600

Db 365 TGATCAGCTGTACACGCGGACGTGCTCGCGGTGATTCGGTCAAGGTTTCGCTCGGTG 424

Qy 601 CGTCCGGTGACCTGACACCGCTTCGCGATATGTGCTCTGCTCCAGGGCGGGGAGACT 660

RESULT 14

US-10-369-493-31202

; Sequence 31202, Application US/10369493

; Publication No. US20030233675A1

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

Db 425 CATCGGCGACCTCGCGCGCTCGCGCATATGTCCGCGCGCTCTCTCGGCGTCCGCGA-- 482

Qy 661 TCTTGGACCGGACCGGACCGCGCTTGAACGCGCGAGAAAGGCTCCGCGCGGACCGGTGC 720

Db 483 -AGTGTTCGCGAAGGGCGAGCGCATGCGCGCCACCGAGGGTCTCGCGCTCTCGCGCTCA 541

Qy 721 AACCGCTCGATCTCTCCATCGGATGCACTGGGCTGGTCAACGGGACCTCCGCCATGA 780

Db 542 AGCGCTCAACGCTGCAAGAACGAGGGGCTGGGCTCTCTGAAACGCGCACGAGGCTTCGA 601

Qy 781 CCGGGATTCGCGCTGGTGAATGCTCACGCTTCGCCCATCTCGGCAACTTGGGCGGTGGCGT 840

Db 602 CGCGCTCGCGCTGTACAAATGTTTCGCCATCGAAGACTGTACCGCACCGCGCTGGTGT 661

Qy 841 TGACGGCCCTGCTTTCGGAATGTCTGAGAGCGCGGACCGAGGCAATGGGCGCGGACTGT 900

Db 662 CGGCGCATTTGTCGCTAGATCGCGCAATGGGCTCGGTCAAGCCGTTTCGACGCGCGATTC 721

Qy 901 CGGACCTCGCGCGCATCCCGGACAGAGACCGCGCAGGAGGCTGGGCGCGCGTGG 960

Db 722 ACAGTTTCGCGCGCATCAAGGTCAGATCGACGCGCGCGGCGATACCGCTCGCTGCTGTG 781

Qy 961 ACGCGACGCGCGGTGCTCGCGCATCTTTCGCGAGCGAGGCTCGACGCGCGCGATA 1020

Db 782 AAGTTTCGCGCAATCAACGTTTCGCGCG----- 808

Qy 1021 TCGGGACGAGCGCGGAGCGCGGAGGATGCTACAGCTTCGCTGCGCTCCCGAGGTTC 1080

Db 809 -----CCGATTGCGACAAAGTGCAGGACCGGTACAGCTTCGCGTCCAGCGCAAGTCA 862

Qy 1081 TCGGGCGGCGCTTCGACACGCTCGCATGCGATGACCGGGTGTGACGATCGAGCTGAACG 1140

Db 863 TGGGCGCGTGTCTGGATCAGATCGCCACCGCGCAACGCTGCTCTCGAAGCGAAGC 922

Qy 1141 CGGTGACCGACAATCCGCTGTTTCGCGCGATGCGAGCGCGCTTCGACCGGGGCA 1200

Db 923 CGGTTCGCGAATTCGCTGATTTTCCCGA---CACCGGGAAGTGTCTGTGGGCGGTA 979

Qy 1201 ATTTATGGGCCAGCATGTGCGCTGACGCTCGATGCGCTGCGACGCGCGCTCAACGCTTC 1260

Db 980 ACTTCCACGAGCGCGTTCGCGCGCGATAACCTTCGCGCTCGCGCGCGCGAAA 1039

Qy 1261 TGGGCGGCTTCGCGAGCGCGCATGTCAGCTGTGACAGATGAAGGCTGNAACGCTGGGC 1320

Db 1040 TCGCGCGCTTGGCGCAACCGCGCATCGCTGCTGTGATCGACGCGACGCTGTGCGGCGCTGC 1099

Qy 1321 TGCCCGCTTCTCCACCGGGCGCGCGCTGTTGAATTCGCGCTTCATGCGGCGCACAGG 1380

Db 1100 CGCC-----GTTCTCTGTCGCGATGCGCGCGTGAATTCGGGCTTCATGATCGCGCACG 1153

Qy 1381 TGACGGCGACCGGCT---CCTGGCGAGATGCGAGCCAGGGACCTGCTTCGATTCATT 1437

Db 1154 TCAGGCTGCGCGCTCGCTTCGGAACAAAGAGCTCGCGCATTCGGCTTCGGTTCGATT 1213

Qy 1438 CGATCTCCAGAACCGCGCAATCAGGATGTGCTGCTTGGGACCATCGCGCGGCGCC 1497

Db 1214 CGCTGCGCACTTCGCGCAACCGAGAACCAACGCTGTGATGCGGAGACGTTTCGCGCGCGCA 1273

Qy 1498 TCTTCCCGGAGAGATCGACCGTTCGGCGGAGATCTTCGCGATCCTCGCTCTCTGCTTG 1557

Db 1274 AGTTCGCGCATTTGCTGAACAAACCGCGGAACATTTCTGTGATCGAACTGCTCGCGCGC 1333

Qy 1558 CACAAGCTCGGAGCTGCGC 1577

Db 1334 CGCAAGCGTCTGATCTGCGC 1353

; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 31202
; LENGTH: 1503
; TYPE: DNA
; ORGANISM: Burkholderia cepacia
US-10-369-493-31202

Query Match 11.4%; Score 211.6; DB 6; Length 1503;
Best Local Similarity 51.3%; Pred. No. 1.4e-45;
Matches 657; Conservative 0; Mismatches 569; Indels 54; Gaps 5;
QY 301 TCGGCGCTGCATCCGAGGCGCGCATGTCTACGACTGACACCGGCTTCGGTCCCC 360
Db 125 TCCTCCGATATCGCGGAGGCGAGCGCGCTTACGGCATCAACCGGGCTTTGGGCGCC 184
QY 361 TTCCGAACCCCTGTATCTCAGGTGAGAAATGTCGAAACGCTGCAGGCAATCTTGTCCATC 420
Db 185 TCGCCAGCAGCATATTCGCACGACCACTGAACTGTGACGCCATCTGGTCTCT 244
QY 421 ATCTGCGACGCGGTGGACCGGTGTTGACTGAGACGACGCGCGCCCATGTTCTCG 480
Db 245 CTCAACGCGTGGCGGTGGTGGAGCGATGTCGGGTCCGGTCTGCTGTCGTCGCGC 304
QY 481 CGCGTCTGGTGTGATGCTCAGGAGACCTCCGGTCCGACGAGGGACCATCGCTCGCC 540
Db 305 TGAACCTCTCGAGCCTCGCGCGCGCCATTCGGGCAATTCGGGTGAAGTGAAGACGCGC 364
QY 541 TGATCAACCTGTCAATTCGAGCTCGCTCCGGCGCTTCCAGCGCGGACGCGTGGCG 600
Db 365 TGATCACTGTATACAGCGAGCTGCTGCGGTGATTCGGTCAAGGTTCCGTGCGTG 424
QY 601 CGTGGGTGACCTGACACCGCTTGGCATGATGCTCTGCTCCAGGCGCGGGAGACT 660
Db 425 CATCGGCGACCTCGCGCGCTCGCATATGTCGGCGCGCTGCTCGCGTCCGGGA-- 482
QY 661 TCCTGACCGGACGAGCGCGCTTGACGGCGACAGGGCTCCGGCGGAGCGCTGC 720
Db 483 -AGTGTTCGGAAGGCGAGCGCATCCGCGCCACCGAGGTTCTCGCGCTCGTCCGCTCA 541
QY 721 AACCGCTCGATCTCTCCCATCGGATGCACTGGCGCTGTCAACGGGACCTCCGCCATGA 780
Db 542 AGCGCTCAGCTGCAAGCAAGAGGGCTGGCGTCTTGAAACGGCAGCGAGCTTCA 601
QY 781 CCGGGATCGCGCTGTGTAATGCTCAGCCTTCGCGCATCTCGGCAACTGGGCGTGGCT 840
Db 602 CGCGCTCGCGCTGTACACATGTTCCGCATCGAAGACCTGTACCGCACCGCGCTGT 661
QY 841 TGACGCCCTCTGTCGGGAATGTTGAGAGCGCGACCGAGGATGGCGCGGCACTGT 900
Db 662 CGGGCGCATTTGCGGTAGATCGCGAATGGGTTCCGTTCAAGCGCTTCGACGCGCGCATTC 721
QY 901 CGCACTCGCGCGCATCCGAGACAGAGCGCGCAGGAGGCTGCGCGCGCGCTGG 960
Db 722 ACGAGTTGCGGCCATCAAGTTCAGATCGACGCGCGCGGCAATCCGCTCGTCTGG 781
QY 961 ACGGACGCGCGGTTGGTCCGCACTGTTCCGAGCGAGGCTTCGACGCGCGGATA 1020
Db 782 AAGTTCCGCAATCAACGTTTCCAGC----- 808
QY 1021 TCGGACGAGCGGAGGCGGGCAGGATGCTACAGCTGCGCTGCGCTCCGAGGTTTC 1080
Db 809 -----CCGATTGCGAAGAGGTGACAGGCCGCTACAGCTGCGCTGCCAGCGCGCATCA 862

QY 1081 TCGGGCGGGCTTCGACACGCTCGCATGCGATGACCGGGTGTGACGATCGAGCTGAACG 1140
Db 863 TGGGCGGCTGTGATCAGATGCGCCACGCGGCCAACTGCTGTCTCGAAGCGAAGC 922
QY 1141 CGGTGACCGACAACTCCGGTGTTCGCGCCGATGAGAGCGTCCCGCTTCGACGCGGGCA 1200
Db 923 CGGTCTCCGACAACTCCGCTGATTTTCCCGA---CACCGCGAAGTGTCTGTGCGGCGGTA 979
QY 1201 ATTTCATGGCGCAGCATGTCGCTGACGTCGATCGCGCTCGCCACGCGGCTCACCGTTC 1260
Db 980 ACTTCACGACAGCGCGTTCGCGCGCTGATTAACCTCGCGCTCGCGCGCGCGGAA 1039
QY 1261 TGGCGGGCTTCGCGAGCGCCAGATTCACAGTGTGACAGATGAAAGGCTGAACCGTGGGC 1320
Db 1040 TCGGCGGCTGCGCGAAGCGCGCATCGCGCTGTGATCGACGCGAGCTGTGCGGCTGC 1099
QY 1321 TGGCCCCCTTCCTCCACCGGGCGCGCGGGTGAATTCGGGCTTCATGGGCGCAAG 1380
Db 1100 CGCC-----GTTCTCGTGCAGTGGCGGCTGAATTCGGGCTTCATGATCGCGCAGC 1153
QY 1381 TGACGCGGACCGGCT---CCTGGCGGATGCGAGCGACCGACCGCTGCTCGATCCATT 1437
Db 1154 TCACGCTCGCGGCTTCGCTTCGGAACAAGACGCTCGCGATCGGCTTCGGTTCGATT 1213
QY 1438 CGATCTCCACGAACGCGCGCAATCAGGATGTGCTTCGCTTCGGGACCATCGCGCGCGCC 1497
Db 1214 CGTGCCTTCGCGCAACCGCGAAGACCGAGTGTGATGCGGAGCTTCGCGCGCGCA 1273
QY 1498 TCTGCGCGAAGATCGACCGTTCGGGCGGAGATCCTTCGATCTTCGCTCTGCTGCTTG 1557
Db 1274 AGCTCGCGACATTCGTGAAAAACACCGCGAACATTCTGTCGATCGAACTGCTCGCGCGC 1333
QY 1558 CACAAGCTGCGAGCTGCGC 1577
Db 1334 CGCAAGGCGTGCATCTGCGC 1353

RESULT 15

US-10-282-122A-13653
; Sequence 13653, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13653
; LENGTH: 1521
; TYPE: DNA
; ORGANISM: Burkholderia fungorum
US-10-282-122A-13653

Query Match 11.4%; Score 211.6; DB 7; Length 1521;
Best Local Similarity 51.3%; Pred. No. 1.4e-45;
Matches 657; Conservative 0; Mismatches 569; Indels 54; Gaps 5;

QY 301 TCGGCGCTGTCATCGCGGAGCGGCGCCATGTCACGAGCTGACACACCGGCTTCGGTCCCC 360
DB 122 TCCCCGATATCGCGCGAAGGGGAGCGCGCCTACGGCATCAACACGGGCTTTGGGCGCC 181
QY 361 TTCGGAACCGCTGATCTCAGGTGAGAACTCTCGAACGCTGCAAGCCCAATCTTGTCCATC 420
DB 182 TCGCCAGCAGCATATTCGCCACGACCAACTCGAACTGTTGACGCGAATCTGGTCTCT 241
QY 421 ATCTGCGCAGCGCGTGGGACCGGTGTTGACTGAGACGAGCGCGCGCCATGGTTCGG 480
DB 242 CTCACCGGTGCGGTGGGTGAGCCGATGTCGGTTCGGTCTGCTGATCGCGC 301
QY 481 CGGCTCTGTTGTCGATCGCTCAGGAGCCTTCGGTTCGACGAGGAGCAATCGCTCGCC 540
DB 302 TGAACCTCTCGAGCCTCGCGCGCGGCCATTCGGGCGAATTCGCGTGAATGATGAGCGCGC 361
QY 541 TGATCGACCTGCTCAATTCGAGCTCGCTCCGCGCTTCCACGCGCGGCACTGGCGG 600
DB 362 TGATCAGCTGTACACGCGCGAGCTGCTCGCGGTGATTCGGTCAAGGGTTCGGTTCGGTG 421
QY 601 CGTTCGGGTGACCTGACACCGCTTTCGCGATATGGTGTCTTCCTCCAGGGCGCGGAGACT 660
DB 422 CATCGGCGACCTCGCGCGCTCGCGCATATGTCGGCGGCGTCTCGGGTCGGCGA-- 479
QY 661 TCTCGACCGGAGCGGACCGCGCTTGAACGCGCGAGAGGGCTTCGGCGCGGACGCGGTGC 720
DB 480 -AGTGTTCGGAAGGGGAGCGATGTCGGGCCACCGAGGGTCTCGGCTCGTTCGGGCTCA 538
QY 721 AACCGCTCGATCTCTCCATCGCATGCACTGGGCTGGTCAACGGGACCTTCGCCATGA 780
DB 539 AGCCGCTCAGCTGCAAGCCAAAGAGGGGCTGGCGCTTCCTGAACCGGACGAGGCTTCGA 598
QY 781 CCGGATCGCGCTGGTGAATGCTCAGCCCTTCGCCCATCTCGGCAACTGGGCGGTGGCGT 840
DB 599 CGGCGCTCGCGCTGTACACATGTTGCCCATCGAAGCTGTACCGCACCGCGCTGGTGT 658
QY 841 TGACGCGCTTCGTTGCGGAATGTCGAGAGCCGGAACCGAGGCAATGGGCGCGGCACTGT 900
DB 659 CGGCGCATTTGTCGGTAGATGCGGCAATGGGCTCGGTCAAGCCGTTTCGACGCGCGCATTC 718
QY 901 CGGACCTGGCGCGCATCCCGACAGAGAGCCCGCAGGAGGCTTCGGCGCGCGGTGG 960
DB 719 ACAGATTGGCGCGCCATCAAGGTCAGATCGACGCGCGCGGCGCATACCGCTCGCTGTGG 778
QY 961 ACGGACGCGCGGCTGGTTCGGGACGCTCATTTGCCGAGCGAGGCTCGACGCGCGCGATA 1020
DB 779 AAGTTTCGGCAATCAACGTTTTCGACG----- 805
QY 1021 TCGGGAACGAGCGGAGCGGCGGAGAGATGCTTACAGCTTCGCTCGCTTCGCAAGTTTC 1080
DB 806 -----CCGATTGCGACAAAGGTGAGGACCCGTTACAGCTTCGCTCGCGTCCAGCGCGCAAGTCA 859
QY 1081 TCGGGGCGGCTTCGACACGCTCGCATGGCATGACCGGGTGTGACGATCGAGCTGAACG 1140

DB 860 TGGGCGGTGTCTGGATCAGATCGGCCACGCGGCCAACGTGCTGCTCGAAGCGAACG 919
QY 1141 CGGTGACCGACAATTCGGGTGTTTCCGCCCGATGGCAGCGTGCCCGCCTTGACACGGGGGCA 1200
DB 920 CGGTCTCCGACAATCCGCTGATTTTCCCCGA---CACCGCGGAAGTGTCTGTCGGGCGGTA 976
QY 1201 ATTTTCATGGGCCAGCATGTGGCGCTGACGTCGATGGCTCGCCACGCGCGCTCACCGTTC 1260
DB 977 ACTTCCACGAGAGCGGTCGCTTCGCGGCCGATTAACCTCGCGCTCGCGCGCGCGGAAA 1036
QY 1261 TGGCGGCGCTTTCGGAGCGCAGATTGCAGCTCTGACAGATGAAGGCTGAAACCGTGGGC 1320
DB 1037 TCGGCGGCTGGCGCAACCGCGCATCGCGTCTGATCGAGCGACGCTGTTCGGGCTGCG 1096
QY 1321 TGCCCCCTTCTCCACCGGGGCCCCCGGGTTGAATTCGGCTTCATGGGCGCACAGG 1380
DB 1097 CGCC-----GTTCTCTGTCGCGGATGGCGCGTGAATTCGGGCTTCATGATCGCGCACG 1150
QY 1381 TGACGGGAGCGCGCT---CCTGGCGGAGATGCGAGCCAGGACCTGCTCGATCCATT 1437
DB 1151 TCACGGCTGCGGCGCTCGCCTCGGAAACAAAGACGCTCGCGCATTCGGCTTCGGTTCGATT 1210
QY 1438 CGATCTCCACGAACCGCGCAATCAGGATGTGCTTCGCTTGGAGCCATCGCGCGCGCC 1497
DB 1211 CGTTCGCCACTTCGGCAACCCAGGAAGACCACTGTCGATGGCGACGTTTCGGCGCGCA 1270
QY 1498 TCTGCGCGAGAGATCGACCGCTTGGGCGGAGATCCTTTCGATCCTTCGCTCTCTGTCTTG 1557
DB 1271 AGCTCGCGACATTTGCTGAAACACCGCGAAACATTTCTGTCGATCGAACTGCTCGCCGCGC 1330
QY 1558 CACAAGCTCGGAGCTGGC 1577
DB 1331 CGCAAGCGCTCGATCTGCGC 1350

Search completed: December 10, 2005, 12:12:48
Job time: 1573.secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 10, 2005, 08:52:45 ; Search time 1129 Seconds

(without alignments)
10920.887 Million cell updates/sec

Title: US-10-621-826-2

Perfect score: 1850

Sequence: 1 ccgaggggccatccacgag.....tcagctgatcccgacgagg 1850

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_21.*
1: geneseqn1980s.*
2: geneseqn1990s.*
3: geneseqn2000s.*
4: geneseqn2001as.*
5: geneseqn2001bs.*
6: geneseqn2002as.*
7: geneseqn2002bs.*
8: geneseqn2003as.*
9: geneseqn2003bs.*
10: geneseqn2003cs.*
11: geneseqn2003ds.*
12: geneseqn2004as.*
13: geneseqn2004bs.*
14: geneseqn2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Length	ID	Description
1	1850	100.0	1850	12	ADJ57135 R. sphaer
2	1446	78.2	1446	13	ADS55829 Bacterial
3	561	30.3	1626	12	ADL16876 Rhodobact
4	244.2	13.2	1587	13	ADS56825 Bacterial
5	242.4	13.1	42000	3	AA63349 Streptomy
6	242.4	13.1	63164	3	AA63348 Streptomy
7	238.4	12.9	1497	13	ADS55996 Bacterial
8	231.6	12.5	1521	8	ACA25972 Prokaryot
9	215.4	11.6	1503	13	ADT4460 Bacterial
10	215.4	11.6	1527	14	ACL71540 M. xanthu
11	215.4	11.6	6439	14	ACL64350 Bacterial
12	211.6	11.4	1503	13	ADS50013 Bacterial
13	211.6	11.4	1503	13	ADS55528 Bacterial
14	211.6	11.4	1521	8	ACA25783 Prokaryot
15	177.4	9.6	1518	13	ADS50000 Bacterial
16	177.4	9.6	1518	13	ADS55516 Bacterial
17	177.2	9.6	1575	13	ADT48701 Bacterial
18	173.4	9.4	1530	4	AA54350 Pseudomon
19	173.4	9.4	1530	8	ACA42730 Prokaryot

20	173.4	9.4	1530	13	ADT42089 Bacterial
21	172.8	9.3	34980	6	ABQ81847 Bifidobac
22	172.8	9.3	34980	6	ABQ81848 Bifidobac
23	171.8	9.3	1548	11	ABD17871 Pseudomon
24	171.8	9.3	2295	11	ABD17522 Pseudomon
25	170.8	9.2	1530	13	ADS59513 Bacterial
26	170.8	9.2	1533	13	ADS62488 Bacterial
27	170.8	9.2	1533	13	ADS63119 Bacterial
28	170.8	9.2	1542	13	ADS60352 Bacterial
29	163.6	8.8	30390	4	AA559520 Propionib
30	163.6	8.8	30390	8	ACF64449 Corynebact
31	163.2	8.8	1527	11	ABD00629 Klebsiell
32	161.8	8.7	1542	6	AA226199 Corynebact
33	161.6	8.7	1521	13	ADS63795 Bacterial
34	161.6	8.7	1521	13	ADS64178 Bacterial
35	161.6	8.7	1521	13	ADS63427 Bacterial
36	159.4	8.6	1566	13	ADS57133 Bacterial
37	157	8.5	1525	8	ACA35512 Prokaryot
38	155.2	8.4	1545	8	ACA45293 Prokaryot
39	153.8	8.3	1536	6	AA26200 Corynebact
40	151.8	8.2	1446	13	ADS50081 Bacterial
41	151.8	8.2	1491	13	ADS55596 Bacterial
42	151.6	8.2	1452	13	ADS56654 Bacterial
43	149.4	8.1	1530	13	ADS61703 Bacterial
44	146.6	7.9	1530	8	ACA44103 Prokaryot
45	144.6	7.8	1461	13	ADS49545 Bacterial

ALIGNMENTS

RESULT 1

ADJ57135
ID ADJ57135 standard; DNA; 1850 BP.

XX AC ADJ57135;

XX DT 06-MAY-2004 (first entry)

XX DE R. sphaeroides tyrosine ammonia lyase (TAL) enzyme encoding DNA.

XX KW TAL; tyrosine ammonia lyase; catalyst; para-hydroxycinnamic acid; PHCA;
XX KW tyrosine; gene; ds.

XX OS Rhodobacter sphaeroides.

XX FH Key Location/Qualifiers

FT CDS

FT 147..1727

FT /*tag= a

FT /product= "TAL"

FT /note= "tyrosine ammonia lyase"

FT /partial

FT /note= "the start codon is not indicated"

XX FN WO2004009795-A2.

XX PD 29-JAN-2004.

XX PF 23-JUL-2003; 2003WO-US023229.

XX PR 23-JUL-2002; 2002US-0397820P.

XX PA (DUPO) DU PONT DE NEMOURS & CO E I.

XX PI Huang L, Xue Z;

xx The invention relates to a tyrosine ammonia lyase enzyme (TAL) and
cc encoding polynucleotides. The TAL polypeptide is used as a catalyst in
cc producing para-hydroxycinnamic acid (PHCA) from tyrosine. The method
cc involves contacting a transformed host cell expressing the polypeptide
cc with a fermentable carbon substrate, e.g. monosaccharides,
cc oligosaccharides, polysaccharides, carbon dioxide, methanol,
cc formaldehyde, formate and carbon-containing amines, preferably glucose,
cc growing the host cell for a time sufficient to produce PHCA, and
cc optionally recovering PHCA. The TAL polypeptide enables simple and cost-
cc effective production of PHCA. The present sequence represents a R.
cc sphaeroides TAL polypeptide encoding DNA.
xx
SQ Sequence 1850 BP; 282 A; 628 C; 635 G; 305 T; 0 U; 0 Other;
Query Match 100.0%; Score 1850; DB 12; Length 1850;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1850; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 CCGAGGCGCCATCCACGAGATGGCGCCGCTTCGGGCTGACGCCCGCGGATCTTCGC 60
Db |||||
Qy 1 CCGAGGCGCCATCCACGAGATGGCGCCGCTTCGGGCTGACGCCCGCGGATCTTCGC 60
Db |||||
Qy 61 TCGATCTGGGCGCGCTGGGCGGACGATCTGAAGCAAGAACTCTGCATTAAACGATTC 120
Db |||||
Qy 61 TCGATCTGGGCGCGCTGGGCGGACGATCTGAAGCAAGAACTCTGCATTAAACGATTC 120
Db |||||
Qy 121 CACTACCTTCGGCGAGAGACAGGAGTGAAGCCAACTGCTGCCATGAGCCCCCGAAGC 180
Db |||||
Qy 121 CACTACCTTCGGCGAGAGACAGGAGTGAAGCCAACTGCTGCCATGAGCCCCCGAAGC 180
Db |||||
Qy 181 CGGCGCTGAGTGGATCGCACATCGATCTGGACAGGCCATGCGTGGCGAGCGCG 240
Db |||||
Qy 181 CGGCGCTGAGTGGATCGCACATCGATCTGGACAGGCCATGCGTGGCGAGCGCG 240
Db |||||
Qy 241 CGCGCGGATTTCTCTTCCCTTCGGCGCGACCGGTCGCTCGAAGCGCGG 300
Db |||||
Qy 241 CGCGCGGATTTCTCTTCCCTTCGGCGCGACCGGTCGCTCGAAGCGCGG 300
Db |||||
Qy 301 TCGGGCTGTATCGCGAGGCGGCCATGTCTACGAGTGAACAACCGGCTCGGTCC 360
Db |||||
Qy 301 TCGGGCTGTATCGCGAGGCGGCCATGTCTACGAGTGAACAACCGGCTCGGTCC 360
Db |||||
Qy 361 TTGCGAACCGCTGATCTCAGTGAGATGTCGAAACGCTGACGCCCAATCTGTTCATC 420
Db |||||
Qy 361 TTGCGAACCGCTGATCTCAGTGAGATGTCGAAACGCTGACGCCCAATCTGTTCATC 420
Db |||||
Qy 421 ATCTGGCAGCGGCTGGGACCGGTGCTTGAATGGAACGACGCGCGGCCATGTTCTGG 480
Db |||||
Qy 421 ATCTGGCAGCGGCTGGGACCGGTGCTTGAATGGAACGACGCGCGGCCATGTTCTGG 480
Db |||||
Qy 481 CGCGTCTGGTTCGATCGCTCAGGAGCGCTCGGTGCGAGCGAGGAGCAATCGCTGCC 540
Db |||||
Qy 481 CGCGTCTGGTTCGATCGCTCAGGAGCGCTCGGTGCGAGCGAGGAGCAATCGCTGCC 540
Db |||||
Qy 541 TGATCGACTCTCAATTCGAGCTCGCTCGCGCGTTTCCAGCGCGGACGCTGGGCG 600
Db |||||
Qy 541 TGATCGACTCTCAATTCGAGCTCGCTCGCGCGTTTCCAGCGCGGACGCTGGGCG 600
Db |||||
Qy 601 CGTGGGCTGACCTGACACCGCTTGGGCAATGTTGCTCTGCTCCAGGGCGGGGAGACT 660
Db |||||
Qy 661 TCCTGGACCGGAGCGGAGCGGCTTGAAGCGGAGAGGCTCGGGCGGAGCGGCTGC 720
Db |||||
Qy 661 TCCTGGACCGGAGCGGAGCGGCTTGAAGCGGAGAGGCTCGGGCGGAGCGGCTGC 720
Db |||||
Qy 721 AACCGCTCGATCTCTCCCATCGATGCACTGGGCTGTCTCAACGAGACCTCCGCCATGA 780
Db |||||
Qy 721 AACCGCTCGATCTCTCCCATCGATGCACTGGGCTGTCTCAACGAGACCTCCGCCATGA 780
Db |||||
Qy 781 CCGGGATGCGCTGGTGAATGCTCACGCTTCCGCCATCTCGGCAACTGGGCGGTGGCGT 840
Db |||||

Db 781 CCGGGATGCGCTGGTGAATGCTCACGCTTCCGCCATCTCGGCAACTGGGCGGTGGCGT 840
Qy |||||
Db 841 TGACGGCCCTGCTTGGCGAATGCTGAGAGCCGAGCCGAGGATGGCGCGGCACTGT 900
Qy |||||
Db 841 TGACGGCCCTGCTTGGCGAATGCTGAGAGCCGAGCCGAGGATGGCGCGGCACTGT 900
Qy |||||
Db 901 CCGACCTTGGCGCGCATCCCGGACAGAAAGACAGCGCGAGGAGGCTGCGCGCCGCGTGG 960
Qy |||||
Db 901 CCGACCTTGGCGCGCATCCCGGACAGAAAGACAGCGCGAGGAGGCTGCGCGCCGCGTGG 960
Qy |||||
Qy 961 ACGGCAGCGCGCGGCTGGTTCGGGACAGTCTATTCGGAGCGGAGGCTGACGCCGCGGAT 1020
Db |||||
Qy 961 ACGGCAGCGCGCGGCTGGTTCGGGACAGTCTATTCGGAGCGGAGGCTGACGCCGCGGAT 1020
Db |||||
Qy 1021 TCGGGACGAGCGCGAGCGGCGGAGGATGCTACAGCTTGGCTTGGCTCCGAGGTTTC 1080
Db |||||
Qy 1021 TCGGGACGAGCGCGAGCGGCGGAGGATGCTACAGCTTGGCTTGGCTCCGAGGTTTC 1080
Db |||||
Qy 1081 TCGGGCGGGCTTCGACACAGCTCGCATGCGATGACCGGGTCTGACGATCGAGCTGAACG 1140
Db |||||
Qy 1081 TCGGGCGGGCTTCGACACAGCTCGCATGCGATGACCGGGTCTGACGATCGAGCTGAACG 1140
Db |||||
Qy 1141 CGGTGACCGCAATCCGGTGTTCGCGCCGAGTGGGACAGGTCGCCCTGACCGGGGCA 1200
Db |||||
Qy 1141 CGGTGACCGCAATCCGGTGTTCGCGCCGAGTGGGACAGGTCGCCCTGACCGGGGCA 1200
Db |||||
Qy 1201 ATTTTCATGGGCGGAGCGGCTGAGCTGCGATGCGCTCGCCACCGCGCTCACCGTTTC 1260
Db |||||
Qy 1201 ATTTTCATGGGCGGAGCGGCTGAGCTGCGATGCGCTCGCCACCGCGCTCACCGTTTC 1260
Db |||||
Qy 1261 TGGCGGCGCTTTCGCGAGCGCGATTCGAGTTCGAGATGAAAGGCTGAAACCGTGGGC 1320
Db |||||
Qy 1261 TGGCGGCGCTTTCGCGAGCGCGATTCGAGTTCGAGATGAAAGGCTGAAACCGTGGGC 1320
Db |||||
Qy 1321 TGCCCCCTTCTTCCACCGGGGCGCGCGGTTGAAATTCGGGTTTCAATGGGCGCACAGG 1380
Db |||||
Qy 1321 TGCCCCCTTCTTCCACCGGGGCGCGCGGTTGAAATTCGGGTTTCAATGGGCGCACAGG 1380
Db |||||
Qy 1381 TGAGGCGAGCGGCTCTTGGCGGAGTGGAGCGGACCGGACCTGCTCGATTCATTCGA 1440
Db |||||
Qy 1381 TGAGGCGAGCGGCTCTTGGCGGAGTGGAGCGGACCGGACCTGCTCGATTCATTCGA 1440
Db |||||
Qy 1441 TCTCCAGAACCGCGCAATCAGGATGTGCTTGGCTTGGGACCATCGCGCGCGCTCT 1500
Db |||||
Qy 1441 TCTCCAGAACCGCGCAATCAGGATGTGCTTGGCTTGGGACCATCGCGCGCGCTCT 1500
Db |||||
Qy 1501 GCGCGAGAAAGATCGACCGTGGGCGGAGATCTTGGCATCTCGCTCTCTGTCTTGAC 1560
Db |||||
Qy 1501 GCGCGAGAAAGATCGACCGTGGGCGGAGATCTTGGCATCTCGCTCTCTGTCTTGAC 1560
Db |||||
Qy 1561 AAGCTGCGGAGCTGCGCTGCGGAGCGGCTAGACGGGGTGTCTCCCGCGGGAGAGC 1620
Db |||||
Qy 1561 AAGCTGCGGAGCTGCGCTGCGGAGCGGCTAGACGGGGTGTCTCCCGCGGGAGAGC 1620
Db |||||
Qy 1621 TGGTGCAGGCGCTTGGCGGAGCGGTTCCCGCGCTTGAAGCGGACCGGCGCTTGGGACG 1680
Db |||||
Qy 1621 TGGTGCAGGCGCTTGGCGGAGCGGTTCCCGCGCTTGAAGCGGACCGGCGCTTGGGACG 1680
Db |||||
Qy 1681 AAATTCGCGCTTGTTCACGACCTTTCGAGCAATCTCCCGTGTGAGCGCGGCTCAGG 1740
Db |||||
Qy 1741 TCGTGGCGGATTCACCGTGAAGGCTTACAGCTCGGATATTTTGGACGAGGCTCTCGATCG 1800
Db |||||
Qy 1741 TCGTGGCGGATTCACCGTGAAGGCTTACAGCTCGGATATTTTGGACGAGGCTCTCGATCG 1800
Db |||||
Qy 1801 ACGCCAGACGCGACGAGACCTTCGATGCGGATCAGCTCGATCCCGACGAGG 1850
Db |||||
Qy 1801 ACGCCAGACGCGACGAGACCTTCGATGCGGATCAGCTCGATCCCGACGAGG 1850
Db |||||

ID AD85829 standard; cDNA; 1446 BP.
XX
AC AD85829;
XX
DT 02-DEC-2004 (first entry)
XX
DE Bacterial polynucleotide #7816.
XX
KW Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cell cycle pathway modification; plant growth regulator;
KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW bacterial polynucleotide; gene; ss.
XX
OS Bacteria.
XX
PN US2003233675-A1.
XX
PD 18-DEC-2003.
XX
PF 20-FEB-2003; 2003US-00369493.
XX
PR 21-FEB-2002; 2002US-0360039P.
XX
PA (CAOY/) CAO Y.
PA (HINK/) HINKLE G J.
PA (SLAT/) SLATER S C.
PA (CHEN/) CHEN X.
PA (GOLD/) GOLDMAN B S.
XX
PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX WPI; 2004-061375/06.
XX
PT New recombinant DNA construct comprising a promoter positioned to provide
PT for expression of a polynucleotide encoding a polypeptide from a
PT microbial source, useful for producing plants with improved properties.
XX
PS Claim 1; SEQ ID NO 31503; 122pp; English.
XX
CC The invention relates to a recombinant DNA construct comprising a
CC promoter functional in a plant cell, where the promoter is positioned to
CC provide for expression of a polynucleotide encoding a polypeptide from a
CC microbial source. The invention also relates to a transformed plant
CC comprising the recombinant DNA construct and a method of producing a
CC transformed plant having an improved property. The plant is a crop plant
CC such as maize or soybean. The method of producing a transformed plant
CC having an improved property comprises transforming a plant with the
CC recombinant DNA construct and growing the transformed plant, where the
CC polynucleotide or polypeptide is useful for improving plant properties.
CC The recombinant DNA construct is useful for producing plants with
CC improved plant properties, e.g. improved cold, heat or drought tolerance,
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC increased resistance to plant disease, better growth rate by modification
CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or
CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition, improved lignin production or improved galactomannan
CC production. This sequence represents a bacterial polynucleotide used in
CC the scope of the invention. Note: The sequence data for this patent did
CC not form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 1446 BP; 212 A; 487 C; 505 G; 242 T; 0 U; 0 Other;
Query Match 78.2%; Score 1446; DB 13; Length 1446;
Best Local Similarity 100.0%; Pred. No. 3.2e-269;
Matches 1446; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 273 GACCGTGCCTGCTCCGAAGCGCGCTCGCGCTGTCTATCCCGAGGCGCGCATGTC 332
DB 1 GACCGTGCCTGCTCCGAAGCGCGCTCGCGCTGTCTATCCCGAGGCGCGCATGTC 60
QY 333 TAGGACTGACAAACCGGCTTCGGTCCCTTTCGAAACCGCCTGATCTCAGGTGAGAAATGC 392
DB 61 TAGGACTGACAAACCGGCTTCGGTCCCTTTCGAAACCGCCTGATCTCAGGTGAGAAATGC 120
QY 393 CGAAGCTGCAGGCGCAATCTTGTCATCATCTGCGCAGCGCGGTGGACCGGTCTTGAC 452
DB 121 CGAAGCTGCAGGCGCAATCTTGTCATCATCTGCGCAGCGCGGTGGACCGGTCTTGAC 180
QY 453 TGGACGACGCGCGCGCATGTTCTGCGCGCTCTGCTGTCGATCGCTCAGGAGCCCTCC 512
DB 181 TGGACGACGCGCGCGCATGTTCTGCGCGCTCTGCTGTCGATCGCTCAGGAGCCCTCC 240
QY 513 GGTGCCAGGAGGCGACCATCGCTCGCTCGATCGACCTGCTCAATCCGAGCTCGCTCCG 572
DB 241 GGTGCCAGGAGGCGACCATCGCTCGCTCGATCGACCTGCTCAATCCGAGCTCGCTCCG 300
QY 573 GCGTTCCAGCGCGCGACCGTGGGCGCTGCGGTGACCTGACACCGCTTGGCATATG 632
DB 301 GCGTTCCAGCGCGCGACCGTGGGCGCTGCGGTGACCTGACACCGCTTGGCATATG 360
QY 633 GTGCTCTGCTCCAGGCGCGGAGACTTCTTGACCGGGACGCGCGCTTGGACGGC 692
DB 361 GTGCTCTGCTCCAGGCGCGGAGACTTCTTGACCGGGACGCGCGCTTGGACGGC 420
QY 693 GCAGAAGGGCTCCGCGCGCGACGCTGCAACCGCTCGATCTCTCCATCGCGATGCACTG 752
DB 421 GCAGAAGGGCTCCGCGCGCGACGCTGCAACCGCTCGATCTCTCCATCGCGATGCACTG 480
QY 753 GCGTGTGTAACGGGACCTCGCCATGACCGGGATCGCGCTGGTGGTGAATGCTCAGCCCTGC 812
DB 481 GCGTGTGTAACGGGACCTCGCCATGACCGGGATCGCGCTGGTGGTGAATGCTCAGCCCTGC 540
QY 813 CGCCATCTCGGCAACTCGGCGGTGCGCTGACGCGCTGCTTGGCGAATGCTCAGAGGC 872
DB 541 CGCCATCTCGGCAACTCGGCGGTGCGCTGACGCGCTGCTTGGCGAATGCTCAGAGGC 600
QY 873 CGGACCGAGGCTCGGCGCGGCACTGTCCGACCTCGCGCGCATCCCGACAGAGGAC 932
DB 601 CGGACCGAGGCTCGGCGCGGCACTGTCCGACCTCGCGCGCATCCCGACAGAGGAC 660
QY 933 GCGCAGCGAGGCTCGCGCGCGCGCTGGAACGCGAGCGCGGGTGGTCCGCGACGTCAAT 992
DB 661 GCGCAGCGAGGCTCGCGCGCGCGCTGGAACGCGAGCGCGGGTGGTCCGCGACGTCAAT 720
QY 993 GCGGAGCGAGGCTCGAGCGCGCGGATATCGGAGCGAGCGCGGGCGGCGAGGATGCC 1052
DB 721 GCGGAGCGAGGCTCGAGCGCGCGGATATCGGAGCGAGCGCGGGCGGCGAGGATGCC 780
QY 1053 TACAGCTCGGCTCGCTCGCGAGTTCCTCGGGCGGGCTTCGACACGCTCGCATGGCAT 1112
DB 781 TACAGCTCGGCTCGCTCGCGAGTTCCTCGGGCGGGCTTCGACACGCTCGCATGGCAT 840
QY 1113 GACCGGTGCTGACGATCGAGCTGAAACCGGTGACCGCAATCCGGTGGTTCGCGCGCAT 1172
DB 841 GACCGGTGCTGACGATCGAGCTGAAACCGGTGACCGCAATCCGGTGGTTCGCGCGCAT 900
QY 1173 GCGAGCTGCGCGCGCTGACCGGGGCAATTCATGCGCGCATGCTGGCGCTGACGCTCC 1232
DB 901 GCGAGCTGCGCGCGCTGACCGGGGCAATTCATGCGCGCATGCTGGCGCTGACGCTCC 960
QY 1233 GATCGCTCGCGACGCGCGCTCACCGTTCCTGCGGGCTTCGCGAGCGCGAGTTCGACGT 1292
DB 961 GATCGCTCGCGACGCGCGCTCACCGTTCCTGCGGGCTTCGCGAGCGCGAGTTCGACGT 1020
QY 1293 CTGACAGTGAAGGCTGAACCGGTGGCTGCCCCCTTCTCTCCACCGGGGCCCCCGGG 1352
DB 1021 CTGACAGTGAAGGCTGAACCGGTGGCTGCCCCCTTCTCTCCACCGGGGCCCCCGGG 1080
QY 1353 TTGAATTCGGCTTCATGGGCGCACAGGTGACGGCGACCGCGCTCTCTGGCGCGAGATCGGA 1412

Db 1081 TTGAATTCGGCTTCATGGCGCACAGGTGACGGCGACCGGCTCTCTGGCCGAGATCGGA 1140
Qy 1413 GCCACGGAGCTGCTCGATCCATTGATCTCCACGAACCGCGCAATCAGGATGTGTC 1472
Db 1141 GCCACGGAGCTGCTCGATCCATTGATCTCCACGAACCGCGCAATCAGGATGTGTC 1200
Qy 1473 TCGCTTGGGACCATCGCGCGGCTCTGCGCGAGAAGATCGACCGTTGGCGGAGATC 1532
Db 1201 TCGCTTGGGACCATCGCGCGGCTCTGCGCGAGAAGATCGACCGTTGGCGGAGATC 1260
Qy 1533 CTTGCGATCTCGCTCTCTGTCTGTGCAAAAGCTCGGAGCTGCGCTCCGCGAGCGCCTA 1592
Db 1261 CTTGCGATCTCGCTCTCTGTCTGTGCAAAAGCTCGGAGCTGCGCTCCGCGAGCGCCTA 1320
Qy 1593 GACGGGTGTCTCCGCGGGGAAGAGCTGTGCGAGCGCCCTGCGCGAGCAGTTCCCGCCG 1652
Db 1321 GACGGGTGTCTCCGCGGGGAAGAGCTGTGCGAGCGCCCTGCGCGAGCAGTTCCCGCCG 1380
Qy 1653 CTTGAGACGACCGCGCCCTGGGACAGGAATTCGCGCGCTTGTACGCACTCTTTCAG 1712
Db 1381 CTTGAGACGACCGCGCCCTGGGACAGGAATTCGCGCGCTTGTACGCACTCTTTCAG 1440
Qy 1713 CAATCT 1718
Db 1441 CAATCT 1446

RESULT 3

ADL16876
ID ADL16876 standard; DNA; 1626 BP.
XX
AC ADL16876;
XX
DT 06-MAY-2004 (first entry)
XX
DE Rhodobacter capsulatus tyrosine ammonia lyase (tal) DNA.
XX
KW Para-hydroxycinnamic acid; photoactive yellow protein;
KW photochromic substance; biochip; liquid crystal polymer; PYP; gene; ds.
XX
OS Rhodobacter capsulatus.
XX
FH Location/Qualifiers
CDS 1..1626
FT /*tag=a
FT /product= "Rhodobacter capsulatus tyrosine ammonia lyase
FT (TAL) protein"
FT /transl_except= (pos:1608..1608, aa:Ser-Ala)

US2004029230-A1.

XX
PD 12-FEB-2004.XX
PF 18-JUN-2003; 2003US-00464609.XX
PR 18-JUN-2002; 2002US-0389593P.XX
PA (KIND/) KYNDT J J A.XX
PA (BEEU/) BEEUMEN J V.XX
PI Kyndt JJA, Beeumen JV;XX
DR WPI; 2004-168893/16.XX
DR P-PSDB; ADL16877.

XX
PT New nucleic acid encoding tyrosine-ammonium lyase polypeptide, for
PT synthesizing para-hydroxycinnamic acid and holo-photoactive yellow
PT protein used as photochromic substance in electro optical random access
PT memory or in biochips.

XX
PS Claim 21; SEQ ID NO 3; 32pp; English.

XX

CC The present invention is generally related to recombinant DNA technology
CC and particularly to DNA strands useful for the production of
CC parahydroxycinnamic acid and photoactive yellow protein (pyp) in a
CC suitable host expression system. The invention is useful for synthesizing
CC parahydroxycinnamic acid and derivative products such as holo-photoactive
CC yellow protein which is useful as a photochromic substance in electro
CC optical random access memory, in biochips, in light-inducible gene
CC expression systems and in liquid crystal polymers. The present sequence
CC is Rhodobacter capsulatus tyrosine ammonia lyase (tal) DNA.
XX
SQ Sequence 1626 BP; 229 A; 528 C; 610 G; 256 T; 0 U; 3 Other;

Query Match 30.3%; Score 561; DB 12; Length 1626;
Best Local Similarity 60.9%; Pred. No. 9.9e-99;
Matches 947; Conservative 0; Mismatches 603; Indels 6; Gaps 2;
Qy 175 CGAAGCGGCGCTGAGCTGGATCGCCACATCGATCTGGACACGAGCCATCGCGTGGCGGA 234
Db 53 CCAAGGACTGCTTCGGCTGGACGGGCGCTGACACTTGTCCAATCGAGGCCATCCCGA 112
Qy 235 GCGGCGGCGCGCGGATTGTCTTGGCCCTCCGCGCGCGACCGGTGCGTGCCTCCGAAG 294
Db 113 CACATCGACCGCGATTTCGGTGACCCCGCGCTGCGGAGCGCTGCGCGCGGCGCATG 172
Qy 295 GCGGCTCGGCGCTGTATCCTCGGAGCGCGCATGTCTACGGACTGACAAACCGGCTTCG 354
Db 173 CCGGCTTGAGCAGCGCATCGCCGAGCAGCGCCACATTTACGGCATCACACCGGCTTCG 232
Qy 355 GTCCCTTGGAAACCGGCTGATCTCAGGTGAGAAATGTCGAAACGCTGAGGCCAAATTTG 414
Db 233 GCCCGCTGGCGAAACCGTCTGATCGGGGCGGATCAGGGGCGGAGCTGCAACGAAACCTGA 292
Qy 415 TCCATCATCTGCGCAGCGGCTGGGACCGGTGCTTGACTGGAGCAGCGCGCGCGCCATGG 474
Db 293 TCTATCATCTGGCCACCGGCTCGGCGCGGAACTGAGCTGGGCGGAGCGCGCGGCTTGA 352
Qy 475 TTCTGGCGGCTGTGTGTGATCGCTCAGGAGCCTCCGCTGCCAGCGAGGGGACCAATCG 534
Db 353 TGTGCGGCGGCTCAACTCGATCCTGCAAGGCGCTCGGGGCGCTCGCCGAGACGATCG 412
Qy 535 CTGCGCTGATCGACCTGCTCAATTCGAGCTCGCTCGGCGGTTCCAGCGCGGCGACCG 594
Db 413 ACCGATCTGCTGCGGTGCTCAATTCGCGGGTTTCCCGCGAGGTTCCGCGCGACGGAACGG 472
Qy 595 TGGGCGGCTCGGCTGACCTGACACCGCTTCGCGCATATGGTGTCTGCTCCAGGCGCGG 654
Db 473 TGGGCGGCTCGGCGGATCTGACCCCGCTTCGCGCATATGCTGCTGGCGCTCGAGGACGG 532
Qy 655 GAGACTTCTTGGACCGGAGCGGACCGGCTTGACGCGCGCAGAAAGGCTCCGCGCGGAC 714
Db 533 GCGCGATGATCGACCCCTCGGGCGCGCTGCGAGAGCGCGGGCGGTGATGGATCGGCTCT 592
Qy 715 GGCTGCAACCGCTCGATCTCTCCATCGGATGATGAGGCTGGGCTGGTCAAGGGACCTCG 774
Db 593 GCGGCGGTCGCTGACGCTGCGCGCGCGTACCGGGCTGGGCGCTGGTGAATGGCACCTCG 652
Qy 775 CCATGACCGGATCGCGCTGGTGAATGCTCACCGCTGCCCGCATCTCGGCAACTGGGCGG 834
Db 653 CGATGACCGGATGCGGCGCTGACCGGGGTGAGGGGCGCGGGCGATCGACCGCGCGC 712
Qy 835 TGGCGTTGACCGGCGCTGCTTTCGGAATGTCTGAGAGCGCGGACCGAGGCTATGGCGCGCG 894
Db 713 TTCGGCACAGCGGCTCTGATGAGGTCTTGTCCGCTCATGCGCGAAGCTTGGCATCCGG 772
Qy 895 CACTGTCCGACTGCGGCGCGCATCCGAGCAGAGGACCGCGCGGAGGCTCGGCGCCC 954
Db 773 CTTTTCGACAGCTGCGCGCGCATCCGCGGCGAGCTGCGGGCGACCGAGCGGCTGCGCGAGG 832
Qy 955 GCGTTCGACGCGCGCGGCTGGTCCGCGACGCTCATTTGCCGAGCGGAGGCTCGACCGCG 1014
Db 833 CGCTGATGGGCGGCGGCTGTCTGCGACCTTGACCGCGCGCGGCTGACCGCGCG 892
Qy 1015 GCGATATCGGAGCGGAGCGCGGCGGAGGATGCTTACAGCCTGCGCTCGGCTCCGCG 1074

893 CGGATCTGGGCCCCGAAGATCATTCGGGCGCAGGATGCTACAGTCTGGCGGTGTGCGGC 952
1075 AGGTTCTCGGGGCGGCTTTCGACACGCTCGCATGATGACCGGGTGTGACGATCGAGC 1134
953 AACTGGTTCGGGCGGCTTTCGGACACGCTCGCATGATGACCGGGTGTGACGATCGAGC 1012
1135 TGAACGGGTGACGACATTCGGTGTTCGCGCGATGCGCGATGCGCGGTGTGACGATCGAGC 1194
1013 TCAATTCCGTCAACGACATTCGATCTTTCCGAGGGTGTGCGGTGTGCGCGGTGTGACG 1072
1195 GGGCAATTTTCATGGGCGCAGATGCGCTCGCATGCGCTCGCCACGCGCGCTCA 1254
1073 GCGCAATTTTCATGGGCGTGCATGTGCGCTTCGATGCGCTCGCATGCGGTGTGACGCGGCGTGG 1132
1255 CCGTTCTCGGCGGCTTTCGGAGCGCCAGATGTCACGCTGTCACAGATGAAGGCTGAAC 1314
1133 TGACGCTGGGCGGCTTTCGAGCGCTGATGCGCGCGCTGACCGACGAAAAGCTGAACA 1192
1315 GTGGGCTGCGGCTTTCCTCCACGCGGCGCGCGCGGCTTGAATTCGGCTTCATGGCG 1374
1193 AGGGCTGCGGCGCTTTCCTGACGCGGCGCGCGCGGCTTGAATTCGGCTTCATGGCG 1252
1375 CACAGGTGACGCGGCGCTTTCCTGCGCGGCGCGGCTTGAATTCGGCTTCATGGCG 1431
1253 GCGAGGTGACGCGGCGCTTTCCTGCGGCGCGCGGCTTGAATTCGGCTTCATGGCG 1312
1432 TCCATTTGATTCACGACGCGCGCTTTCCTGCGGCGCGCGGCTTGAATTCGGCTTCATGGCG 1491
1313 TGCAGTCTGCTCGACCAATGGCGCATGAGATGTGCTCGATGGAACGATGCGG 1372
1492 CGGCGCTTTCGCGGCGGAGATGACGCTTTCGCGGCGGAGATCTTCGATCTTCGCTCT 1551
1373 CGGCGAGGCGCGGCGGCGGCTTTCGCGGCGGAGATCTTCGATCTTCGCTCTTCGCTCT 1432
1552 GTCTTGACAGCTGCGGAGCTGCGCTGCGGCGGAGCTTTCGATCTTCGCTCTTCGCTCT 1608
1433 CCTTGTGCGGCGGATGATCTGCTTGCAGACCGCGGCGGAGGCGGCGGAGTGTGCTTGA 1492
1609 CGGCGAAGAGCTGCTGAGCGCTTTCGCGGCGGAGCTTTCGCGGCGGAGGCGGAGGCGG 1668
1493 CGGCGGCGGATCTGCGGCGGAGCTGCGGCGGAGCTTTCGCGGCGGAGGCGGAGGCGG 1552
1669 CCTTGGGAGAGAAATTCGCGGCTTTCGATGCGGCTTTCGAGGAGTTCGCGGCT 1724
1553 CGCTTGGCGGCGGATGATGAGCGGCTTTCGAGGCTTTCGAGGCTTTCGAGGCTTTCGAGG 1608

RESULT 4

AD56825
ID AD56825 standard; cDNA; 1587 BP.
XX AC AD56825;
XX DT 02-DEC-2004 (first entry)
XX DE Bacterial polynucleotide #8812.
XX KW Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cell cycle pathway modification; plant growth regulator;
KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW bacterial polynucleotide; gene; ss.
XX OS Bacteria.
XX PN US2003233675-A1.
XX PD 18-DEC-2003.
XX PF 20-FEB-2003; 2003US-00369493.

21-FEB-2002; 2002US-0360039P.
(CAOY/) CAO Y.
(HINK/) HINKLE G J.
(SLAT/) SLATER S C.
(CHEN/) CHEN X.
(GOLD/) GOLDMAN B S.
Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
WPI; 2004-061375/06.
New recombinant DNA construct comprising a promoter positioned to provide
for expression of a polynucleotide encoding a polypeptide from a
microbial source, useful for producing plants with improved properties.
Claim 1; SEQ ID NO 32499; 122pp; English.
The invention relates to a recombinant DNA construct comprising a
promoter functional in a plant cell, where the promoter is positioned to
provide for expression of a polynucleotide encoding a polypeptide from a
microbial source. The invention also relates to a transformed plant
comprising the recombinant DNA construct and a method of producing a
transformed plant having an improved property. The plant is a crop plant
such as maize or soybean. The method of producing a transformed plant
having an improved property comprises transforming a plant with the
recombinant DNA construct and growing the transformed plant, where the
polynucleotide or polypeptide is useful for improving plant properties.
The recombinant DNA construct is useful for producing plants with
improved plant properties, e.g. improved cold, heat or drought tolerance,
tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
increased resistance to plant disease, better growth rate by modification
of the cell cycle pathway with plant growth regulators, increased rate of
homologous recombination, modified seed oil or protein yield and/or
content, improved yield by modification of carbohydrate, nitrogen or
phosphorus use and/or uptake, by modification of photosynthesis or by
providing improved lignin production or improved galactomannan
condition, improved lignin production or improved galactomannan
production. This sequence represents a bacterial polynucleotide used in
the scope of the invention. Note: The sequence data for this patent did
not form part of the printed specification but was obtained in electronic
format from USPTO at seqdata.uspto.gov/sequence.html.
Sequence 1587 BP; 261 A; 522 C; 539 G; 265 T; 0 U; 0 Other;
Query Match 13.2%; Score 244.2; DB 13; Length 1587;
Best Local Similarity 50.2%; Pred. No. 1.1e-37;
Matches 780; Conservative 0; Mismatches 743; Indels 30; Gaps 6;
Qy 178 AGCGGCGGCTGAGCTGGATCGCCACATCGATCTGGACACGCGCCATCGCGTGGCGAGCG 237
Db 2 ATCGGCGGACATCGACGGCCACCATCTGACCCCGACACCGTGGCGCCATCGCACGCG 61
Qy 238 GCGGCGCGGGATGTCTTTCCTTCCCTTCGGGCGCGGACCGGTCGCGTCCGAGCGG 297
Db 62 GCCAGCGCGCGCCATCGTCTCCGAGCGCGCTCTCGGCAAGGTTGCGCGATGCGCGCGCC 121
Qy 298 GGCTCGCGGCTGTTCATCGGCGGCGCGCATGTCTACGACTGACACCGGCTTCGGTC 357
Db 122 GTTTCGAGAGGTGGTGGCGGCCAATGTGCGGATGTACGGCGTCTTCACGGGCTTTGGCG 181
Qy 358 CCCTTGGGAACCGCTCATGTCTCAGGTGAGATGTCCGAACGCTGACGCGCAATCTTGTTC 417
Db 182 AACTGTGACACACTGGGTGACATCGAATGCGCGTGGCTGACGAGACCTGTTGC 241
Qy 418 ATCATCTGGCCAGCGGCTGGGACCGGTGTGATGGAACGCGCGCGCCCATGTTTC 477
Db 242 GCAGCCATTGCGGCGGTGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 301
Qy 478 TGGGCGCTGTGGTGTGATCGCTCAGGAGGCTTCGGTGGCGGAGGAGGAGGAGGAGGAGG 537
Db 302 TCGCGCGTGCCTTGGCGAGGAGTACTCGGCGGTGCGCGCGGCGGCGGCGGCGGCGGCGG 361

QY 538 GCCTGATCGACCTGCTCAATTCCGAGCTCGCTCCGGCCGTTCCAGCCGCGGCAACGTTGG 597
Db 362 AACTGCTGAAGTATCTGAAAGCCGGCATCACGCCGCTGCCAGAGTGGTTGCTCG 421
QY 598 GCGCGTCGGGTGACCTGACACCGCTTCGGCATATGGTCTCTGCTCCAGGCGCGGGAG 657
Db 422 GGGCCAGCGGTGATCTCGCGCTCTGTCGACGTCGCGCATCACGCTGATCGCGGAAGCA 481
QY 658 ACTTCTCGGACCGGACGGACGGGCTTGACGGCGCAGAGGCTCCGCGCGGACGGC 717
Db 482 AGTGTCTGACCGAGATGGCGTACGGACCCACGGCCGAAGTCTCTCGCGAGGTGGCA 541
QY 718 TGCACCGCTCGATCTCTCCCATCGCGATGCACTGGCGCTGGTCAACGGGACCTTCGCCCA 777
Db 542 TCACCGCGCTCGCGCTGCTACAAGGAAGGCTGGCGCTGATCAACGGGACATCGGCCCA 601
QY 778 TGACCGGGATCGCGCTGGTGAATGCTCACGCTGCGCCCATCTCGGCAACTGGGCGGTGG 837
Db 602 TGACCGGGGTGCTGCTGCTGCTGAGAGCGCTGCGCGCGCAGGTCCAGCAGGCGCGAGA 661
QY 838 CGTTGACGGCCCTGCTTTCGGGATGCTGAGAGCGCGACCGAGGCAT---GGCGCGCG 894
Db 662 TCATCGCGCGCTGGCGCTCGAAGACTATCCGCTCGCGCGATGCTTTCATGGCCCATG 721
QY 895 CACTGTCCGACCTGCGCGCCGATCCCGGACAGAGGACGCGCGAGCGAGGCTCGCGCCCC 954
Db 722 GGCACGACATCGCCAAACCGCATCCGGGACAGATCCGCTCGCGCGGCAACATGGCGCGC 781
QY 955 GCGTGAACGGACGCGCGGGGTGTCGGGCAG-----TCATTGCCGAGCGGAGGTCTG 1008
Db 782 TGCTGGCGGATTCGGCAACGGCTCTCCGGACATGGCGAACTGTCCGCGAGATGAAGACAC 841
QY 1009 ACGCGCGGATATCGGACGAGCGCGAGCGGG-----CAGGATGCTCAGACTGC 1062
Db 842 GCGCGGCGAGGCGCAAGAACACCGGCACTGGCGTGTTCATCCAGAGGCTCACACGCTGC 901
QY 1063 GCTGCGCTCGCGAGTTCTCGGGCGGGCTTCGACACGCTCGCATGCAATGACCGGTGC 1122
Db 902 GCTGCATTCGCGAGTCTTGGCGGTGGCGATAGCTCGACCATTTGCGCACCGTGG 961
QY 1123 TGACGATCGAGCTGAACGCGGTGACCGACAATCGGTGTTTCCGCCCGATGGCAGCGTGC 1182
Db 962 TCGAGCGCAACTGAATTCATCGAATGACAATCGCTGTTTCTTCAAGACGCGGAGCTGT 1021
QY 1183 CCGCCCTGACGGGGCAATTTCATGGCCAGCATGTGGCGCTGAGCTCCGATCGCGCTCG 1242
Db 1022 TC-----CAGCGCGGCAACTTCACGGCCAGCAGGTGGCATTCGCAATGGACTTCTCTGG 1075
QY 1243 CCACGGCGCTCACCGTTCTGGCGGGCTTCGGAGCGCCAGATTGCACTCTGACAGATG 1302
Db 1076 CCATCGCGCCACGCAACTGGCGGTGGTGTGCGAGCGCCGCTGACCGCTCTCTGAGCC 1135
QY 1303 AAGGCTGAACCGTGGGCTGCCGCCCTTCTTCACCGGGGCCCGCGGGTTGAAATTCG 1362
Db 1136 CGCATCTGAACAACAATCTCGCGCGTTCCTGGCGCGGCGAAGCAGAGGGTGTGCTGTCG 1195
QY 1363 GCTTCATGGGGCGACAGGTGACCGCGACCGGCTCTCGCGGAGATGCGAGC---CACGG 1419
Db 1196 GGTTTCGGGGGACAGATATCCGGCCACGCGGTGATTGCGCGAGAACCGCACCATTCGA 1255
QY 1420 GACTGCTCTGATTCGATTCGATCTCAACGAACGCGCAATCAGGATGTGGTCTCGCTTG 1479
Db 1256 GCCCGGCGAGCATCCAGAGTGTGCGCTCGAACCGCGGACACAGGATGTGTGAGATGG 1315
QY 1480 GGAACATCGCGCGCGCTCTCGCGGAGAAATCGACCGTTGGGGCGGAGATCTTTGCGCA 1539
Db 1316 GGCTGATCGCTCGCGCAATCGCGCATTTCTCGACATAACCACTAGTACATCTCTCGCGC 1375
QY 1540 TCCTCGCTCTCTGTCTTGACACAGCTGCGAGCTCGCTCGCGCAGCGGCTAGACGGGG 1599
Db 1376 TGGAGTTGCTGGGCTCATGTGAGCGCGCCGAACCT-----CGCGGGCGCGGTGAGCAAC 1429

QY 1600 TGTCTCCCGCGGGAGAAAGCTGGTGCAGGCCCTGCGCGAGCAGTTCCCGCCGCTTGAGA 1659
Db 1430 TGGCGCGCGCAGCGCGCGCTGTTGCGCTTCGTGCGGAGCGCTGCGCTTCTCTGCGA 1489
QY 1660 CGGACGCGCCCTGGGACAGGAATTGCGCGCTTGTCTAGCAGCCTCTTGCAG 1712
Db 1490 TCATCGCTATATGACCGACGACATCAGGCTATGCGCGCGCTGCTCCGTCAG 1542

RESULT 5
AA863349/C
ID AA863349 standard; DNA; 42000 BP.
XX
AC AA863349;
XX
DT 06-MAR-2001 (first entry)
XX
DE Streptomyces globisporus C-1027 gene cluster ORF -7-25.
KW Eneidyne C-1027 biosynthesis gene cluster; apoprotein; chromophore;
XX cancer; ds.
XX
OS Streptomyces globisporus.
XX
FH Key
CDS Location/Qualifiers
FT complement (8..658)
FT /*tag= a
FT /product= "ORF -7 protein"
FT complement (930..1478)
FT /*tag= b
FT /product= "ORF -6 protein"
FT complement (1649..2713)
FT /*tag= c
FT /product= "ORF -5 protein"
FT complement (2850..3237)
FT /*tag= d
FT /product= "ORF -4 protein"
FT complement (3442..4971)
FT /*tag= e
FT /product= "ORF -3 protein"
FT 5982..7479
FT /*tag= f
FT /product= "glycerol phosphate transporter"
FT complement (7573..9900)
FT /*tag= g
FT /product= "ABC transport/UvrA-like protein"
FT complement (9982..11349)
FT /*tag= h
FT /product= "Na+/H+ transporter"
FT complement (11351..12835)
FT /*tag= i
FT /product= "hydroxylase/halogenase"
FT 13012..14079
FT /*tag= j
FT /product= "dNDP-glucose synthase"
FT complement (14212..14643)
FT /*tag= k
FT /product= "CagA"
FT complement (14690..15922)
FT /*tag= l
FT /product= "aminotransferase"
FT complement (15919..16653)
FT /*tag= m
FT /product= "N-methyl transferase"
FT complement (16653..17924)
FT /*tag= n
FT /product= "C-methyl transferase"
FT complement (18031..19191)
FT /*tag= o
FT /product= "epoxide hydase"
FT complement (19267..19929)
FT /*tag= p
FT /product= "anthranilate synthase II"

FT CDS complement (19926. .21407)
 FT /*tag= q
 FT /product= "anthranilate synthase I"
 FT complement (21424. .22878)
 FT /*tag= r
 FT /product= "coenzyme F390 synthetase"
 FT complement (22875. .23546)
 FT /*tag= s
 FT /product= "iron-sulphur flavoprotein"
 FT complement (23566. .24702)
 FT /*tag= t
 FT /product= "O-acyl transferase"
 FT complement (24986. .25564)
 FT /*tag= u
 FT /product= "epimerase"
 FT 25815. .27170
 FT /*tag= v
 FT /product= "monooxygenase"
 FT 27214. .28593
 FT /*tag= w
 FT /product= "glycosyl transferase"
 FT 28590. .29588
 FT /*tag= x
 FT /product= "dNDP glucose dehydratase"
 FT 29632. .31197
 FT /*tag= y
 FT /product= "SgCB transmembrane efflux protein"
 FT 31280. .32590
 FT /*tag= z
 FT /product= "coenzyme F390 synthetase"
 FT 32809. .34392
 FT /*tag= aa
 FT /product= "hydroxylase"
 FT complement (34458. .35294)
 FT /*tag= ab
 FT /product= "ORF 22 protein"
 FT /transl_except= (pos:35223. .35225,aa:Ile)
 FT /transl_except= (pos:35226. .35228,aa:Ala)
 FT complement (35518. .35938)
 FT /*tag= ac
 FT /product= "ORF 18 protein"
 FT 36165. .37490
 FT /*tag= ad
 FT /product= "coenzyme F390 synthetase"
 FT 37559. .38938
 FT /*tag= ae
 FT /product= "type II NRPS condensation enzyme"
 FT 38983. .39264
 FT /*tag= af
 FT /product= "type II peptidyl carrier protein"
 FT complement (39367. .40986)
 FT /*tag= ag
 FT /product= "aminomutase"
 FT complement (41052. .42000)
 FT /*tag= ah
 FT /product= "type II NRPS adenylation enzyme"
 FT /partial
 FT WO200040596-A1.
 FT PN
 FT XX
 PD 13-JUL-2000.
 XX
 PP 06-JAN-2000; 2000WO-US000446.
 XX
 PR 06-JAN-1999; 99US-0115434P.
 PR 05-JAN-2000; 2000US-00477962.
 XX
 PA (REGC) UNIV CALIFORNIA.
 XX
 XX Shen B, Liu W, Christenson SD, Standage S;
 XX WPI: 2000-465947/40.
 DR P-PSDB; AAB13554, AAB13555, AAB13556, AAB13557, AAB13558, AAB13559,

DR AAB13560, AAB13561, AAB13562, AAB13563, AAB13564, AAB13565, AAB13566,
 DR AAB13567, AAB13568, AAB13569, AAB13570, AAB13571, AAB13572, AAB13573,
 DR AAB13574, AAB13575, AAB13576, AAB13577, AAB13578, AAB13579, AAB13580,
 DR AAB13581, AAB13582, AAB13583, AAB13584, AAB13585, AAB13586, AAB13587.
 XX
 PT Isolated nucleic acid comprising a nucleic acid encoding any of C-1027
 PT open reading frames (ORFs) -7 to 42, excluding ORF 9 (cagA), useful for
 PT the production of enediyne C-1027 antitumor antibiotics.
 XX
 PS Claim 1; Page 78-130; 160pp; English.
 XX
 CC The present sequence is the first 42000 bases of the enediyne C-1027 gene
 CC cluster from Streptomyces globisporus. Enediyne C-1027 is an antibiotic,
 CC consisting of an apoprotein and a non-peptidic chromophore, which acts by
 CC damaging DNA. The sequences within the gene cluster, and the proteins
 CC they encode, can be used in the treatment of cancer, along with
 CC antagonists of the protein. Each of the open reading frames is
 CC specifically claimed, excluding ORF 9, which encodes CagA
 XX
 SQ Sequence 42000 BP; 6282 A; 14213 C; 15007 G; 6498 T; 0 U; 0 Other;
 Query Match 13.1%; Score 242.4; DB 3; Length 42000;
 Best Local Similarity 50.3%; Pred. No. 2.7e-37;
 Matches 749; Conservative 0; Mismatches 706; Indels 33; Gaps 5;
 Qy 227 CGTGGCAGCGGCGCGCGGATTGCTTCCCTCCGCGCGCGGACCGGTGCGGTGC 286
 Db |||||
 Qy 40906 CGTGGCAGCGGCGCGCGGATTGCTTCCCTCCGCGCGCGGACCGGTGCGGTGC 40847
 Db |||||
 Qy 287 GTCGGAAGCGCGCTCGCGCTGTCATCCGCGAGCGCGCATCTCTACGAGTGCAGAAC 346
 Db |||||
 Qy 40846 GAGCGGAGATCTTCGAGGGATCGCGGAACAGAACATCCCCATCTACGGGTGACAC 40787
 Db |||||
 Qy 347 CGGCTTCGGTCCCTTCGGAACCGCTGTCTCAGGTGAGAAATGTCGGAACGTGCAGGC 406
 Db |||||
 Qy 40786 CGGTCACGCGAGATGATCTACATGCAAGTGCAGAGGAGTTCGAACTGCAGAC 40727
 Db |||||
 Qy 407 CAATCTTGTTCATCATCTGGCCAGCGCGGTGGGACCGGTGCTTACCTGGACGACGCGCG 466
 Db |||||
 Qy 40726 CAATCTCGTCGTAGCCACAGCGCGGAGTGGTCCGCTGTTCCCGGAGGACGAGGCGCG 40667
 Db |||||
 Qy 467 CGCATGGTTCGGCGGTCTGGTGTGATCGCTCAGGAGGACCTCGGTGCCAGGAGGG 526
 Db |||||
 Qy 40666 GGGGATCGTCGCGCGCGGTGAAACACCTTCGCAAGGGCCACTCCCGGTGCGCCCAT 40607
 Db |||||
 Qy 527 GACCATCGCTCGCTGATCGACCTGCTCAATTCGAGCTCGCTCCGCGCGTTCCTCAGCG 586
 Db |||||
 Qy 40606 CATCTCGAAGGCTCGCGCAGTACCTGACGAGGGCATCACCCGCGCATACCGAGAT 40547
 Db |||||
 Qy 587 CGGACCGTGGCGCGTGGGTGACCTGACACCGCTTGGCGATATGCTGTGCTGCTCCA 646
 Db |||||
 Qy 40546 CGGTCACATCGGGCGAGCGCGACCTGGCTCCCTCTCCACATCGTCGAGCACCTCAT 40487
 Db |||||
 Qy 647 GGGCGGGAGACTTCTCGACCGGAGCGGAGCGGCTTGCAGCGGCAGAGAGGCTCCG 706
 Db |||||
 Qy 40486 CGGAGGGGTAGCTGCTG---CGCGACGACGGCGGTGGAGACCGCCCGGCTGCGC 40430
 Db |||||
 Qy 707 GCGCGACGGCTGCAACCGCTCGATCTCTCCATCGCGATGCTGCGGTGCTGCTCAACGG 766
 Db |||||
 Qy 40429 CGAGCGGGCATCGAGCGGCTCGAATCGCTTCAGGAGGGCTCCGACTGATCAACGG 40370
 Db |||||
 Qy 767 GACCTCCGCATGACCGGGATCGCGGTGGTGAATGCTACGCTTGCCTGCGCCATCTTCGGA 826
 Db |||||
 Qy 40369 CAGTCCGGGATGACCGGTCTGGGCTCCCTGGTCTCGGACGGGCTTCGGAGGAGGCCA 40310
 Db |||||
 Qy 827 CTGGCGGTGGCTTGACCGGCTGCTTGGGAATGCTCTGAGAGGCGCGGACCGAGGATG 886
 Db |||||
 Qy 40309 GCAGGCGGAGATCGTGACGGCTCTGCTCATCGAGGGGTGACGGGATCGACCCCTT 40250
 Db |||||
 Qy 887 GGGCGCGGACTGTCCGACCT---GCGCGCGCATCCCGGACAGAGGACCGCGCAGCGAG 943
 Db |||||
 Qy 40249 CCTCGCGGAGGGGACGACATAGCCCGCCCGCAGGAGGCGAGATCGACACCGCGGCAA 40190
 Db |||||

Qy	944	GCTCGCGCCGCGGTGTGACCGGACGCGCGCGGTGGTTCGGGACGCTCATTTGCCGAGCGGAG	1003	FT	/*tag= b
Db	40189	CATCGCGGCCCTGATGCGGGGCGCAGCGACTGACGCTCGAGCAGCCGACCTCGCGCCGAGA	40130	FT	/product= "ORF -6 protein"
Qy	1004	GCTC-----GACGCGCGGATATCGGGACGGAGCGGAGCGGGGCGGAGGA	1048	FT	complement (1649. .2713)
Db	40129	ACTCCAGAAGACAAAGAGCGCCCGCAAGGACGCTCCAGCGCTCGGAGATCTACTCTGCAGAA	40070	FT	/*tag= c
Qy	1049	TGCCTACAGCTCGCTCGCTCCGACGAGTTCTCGGGGGGGCTTCGACACGCTCCGATG	1108	FT	/product= "ORF -5 protein"
Db	40069	GGCTACTCTGCGGGCCATCCCGACGTCGTCGGGCGGTGCGCGACACCTTGTACCA	40010	FT	complement (2850. .3237)
Qy	1109	GCATGACCGGGTGTGACGATCGAGCTGAACCGGTCACCGACAATCCGTTTCGCGC	1168	FT	/*tag= d
Db	40009	CGCGCGGCACAAAGCTCGGCATCGAGCTCAACTCGGCCAACACAAACCGCTCTT-----	39956	FT	/product= "ORF -4 protein"
Qy	1169	CGATGGCAGCTGCCCGCTCGACGGGGGCAATTTATGGCCAGCATGTGGCGGTGAC	1228	FT	complement (3442. .4971)
Db	39855	CTTCGAGGGCAAGGAGATCTTCCACGGGGCGAACTTCCACGGTCAGCCGATCGCGTTCCG	39896	FT	/*tag= e
Qy	1229	GTCGATGCGCTCGCCACGGCGCTCACCGTTCTGGCGGGCTTCGCGAGCGCCAGATTGC	1288	FT	/product= "ORF -3 protein"
Db	39895	GATGGACTTCTGTGACATCGCGCTCACCCAGCTCGGCGTCTGGCCGAGCGGACGATCAA	39836	FT	5982. .7479
Qy	1289	ACGCTGTGACAGATGAAGGCTGAACCGTGGGTGCCCGCTTCCTCCACCGGGGGCCCCG	1348	FT	/*tag= f
Db	39835	CCGGTCTCTGAACCGGCACCTCAGCTACGGCTCCCGGAGTTCTCTGTCGCGGACCC	39776	FT	/product= "glycerol phosphate transporter"
Qy	1349	CGGTTGAATTCGGCTTCATGGGGCGCACAGTGACGGGACCGCGCTCTCTGCCGAGAT	1408	FT	complement (7573. .9900)
Db	39775	GGGCTGTGACAGCGGATTCGCGCGGCCCACTACCCGGCCACCGCACTGGTGGCCGAGAA	39716	FT	/*tag= g
Qy	1409	GCAGGCCACGGGACCTGCTCTCGATCCATTGATCTCCACGAACGCGCCCAATCAGGATCT	1468	FT	/product= "ABC transport/UvrA-like protein"
Db	39715	CCGACATCGCGCCCGCCAGCACCCAGAGCTCCCGTCCACGGCGACACCAAGACGT	39656	FT	complement (9982. .11349)
Qy	1469	GGTCTCGCTGGGACCATCGCGCGCGCTCTCGCGCGAGAAAGATCAGCGTTGGGCGGA	1528	FT	/*tag= h
Db	39655	GGTGAGCATGGGCTGATCTCGCGCCGCAAGCGCCGCGGTCTCTGCGAACACAAACAA	39596	FT	/product= "Na+/H+ transporter"
Qy	1529	GATCCTTGCGATCTCGCTCTCTGTGTCGAAGCTGCGGAGTGGCTGCGCGGACGCGG	1588	FT	complement (11351. .12835)
Db	39595	GATCCTCGCG-----GTGGAGTACCTTGGCGCGCCGCGCGGTGCACATCTCCGCGCG	39542	FT	/*tag= i
Qy	1589	CCTAGACGGGTGTCCTCCGCGGGAGAAAGCTGTGTCAGGCCCTGCGCGAGCAGTTCCC	1648	FT	/product= "hydroxylase/halogenase"
Db	39541	GTTGACGGCTTAGCCCCGCGCGGAAGGCCACGTACGAAGCGGTGCGCGCTGTTTCC	39482	FT	13012. .14079
Qy	1649	GCGCTTGAGACGGACCGGCCCTTGGGACAGGAAATTCGCGCGCTTGC	1696	FT	/product= "dNDP-glucose synthase"
Db	39481	GACGCTGGCGTCGACCGGTACATGCGCGACGACATCGAGCTGGTCCG	39434	FT	complement (14212. .14643)
RESULT 6					
AAA63348/c					
ID	AAA63348 standard; DNA; 63164 BP.				
XX					
AC	AAA63348;				
XX					
DT	06-MAR-2001 (first entry)				
DE	Streptomyces globisporus C-1027 gene cluster.				
XX					
KW	Eneidine C-1027 biosynthesis gene cluster; apoprotein; chromophore;				
KW	cancer; ds.				
OS	Streptomyces globisporus.				
XX					
PH	Location/Qualifiers				
FT	complement (8. .658)				
FT	/*tag= a				
FT	/product= "ORF -7 protein"				
FT	complement (930. .1478)				

```
FT CDS /product= "coenzyme F390 synthetase"
FT 32809..34392
FT /tag= aa
FT /product= "hydroxylase"
FT complement(34458..35294)
FT /tag= ab
FT /product= "ORF 22 protein"
FT /transl_except= (pos:35223..35225,aa:ile)
FT /transl_except= (pos:35226..35228,aa:ala)
FT complement(35518..35938)
FT /tag= ac
FT /product= "ORF 18 protein"
FT 36165..37490
FT /tag= ad
FT /product= "coenzyme F390 synthetase"
FT 37559..38938
FT /tag= ae
FT /product= "type II NRPS condensation enzyme"
FT 38983..39264
FT /tag= af
FT /product= "type II peptidyl carrier protein"
FT complement(39367..40986)
FT /tag= ag
FT /product= "aminomutase"
FT complement(41052..42611)
FT /tag= ah
FT /product= "type II NRPS adenylation enzyme"
FT 43945..46024
FT /tag= ai
FT /product= "transmembrane transport protein"
FT 46167..47171
FT /tag= aj
FT /product= "O-methyl transferase"
FT 47228..48484
FT /tag= ak
FT /product= "p450 hydroxylase"
FT complement(48607..49714)
FT /tag= al
FT /product= "oxidoreductase"
FT 50350..51390
FT /tag= am
FT /product= "ORF 31 protein"
FT complement(51421..52341)
FT /tag= an
FT /product= "oxidoreductase"
FT 53241..54074
FT /tag= ao
FT /product= "ORF 33 protein"
FT 54231..55379
FT /tag= ap
FT /product= "ORF 34 protein"
FT 56026..56880
FT /tag= aq
FT /product= "proline oxidase"
FT complement(56927..57736)
FT /tag= ar
FT /product= "ORF 36 protein"
FT complement(57833..58304)
FT /tag= as
FT /product= "ORF 37 protein"
FT complement(58440..60095)
FT /tag= at
FT /product= "p450 hydroxylase"
FT complement(60092..60621)
FT /tag= au
FT /product= "ORF 39 protein"
FT 60940..62019
FT /tag= av
FT /product= "ORF 40 protein"
FT 62045..62899
FT /tag= aw
FT /product= "ORF 41 protein"
FT complement(62787..63164)
```

```
FT FT /*tag= ax
XX /product= "ORF 42 protein"
XX WO200040596-A1.
XX 13-JUL-2000.
XX
XX 06-JAN-2000; 2000WO-US000446.
XX
XX 06-JAN-1999; 99US-0115434P.
XX
XX 05-JAN-2000; 2000US-00477962.
XX
XX (RSGC ) UNIV CALIFORNIA.
XX
XX Shen B, Liu W, Christenson SD, Standage S;
XX
XX WPI; 2000-465947/40.
XX
XX P-PSDB; AAB13554, AAB13555, AAB13556, AAB13557, AAB13558, AAB13559,
XX AAB13560, AAB13561, AAB13562, AAB13563, AAB13564, AAB13565, AAB13566,
XX AAB13567, AAB13568, AAB13569, AAB13570, AAB13571, AAB13572, AAB13573,
XX AAB13574, AAB13575, AAB13576, AAB13577, AAB13578, AAB13579, AAB13580,
XX AAB13581, AAB13582, AAB13583, AAB13584, AAB13585, AAB13586, AAB13587,
XX AAB13588, AAB13589, AAB13590, AAB13591, AAB13592, AAB13593, AAB13594,
XX AAB13595, AAB13596, AAB13597, AAB13598, AAB13600, AAB13601, AAB13602,
XX AAB13603, AAB13606, AAB13607.
XX
XX Isolated nucleic acid comprising a nucleic acid encoding any of C-1027
XX open reading frames (ORFs) -7 to 42, excluding ORF 9 (caga), useful for
XX the production of enediyne C-1027 antitumor antibiotics.
XX
XX Claim 1; Page 78-157; 160pp; English.
XX
XX The present sequence is the enediyne C-1027 gene cluster from
XX Streptomyces globisporus. Enediyne C-1027 is an antibiotic, consisting of
XX an apoprotein and a non-peptidic chromophore, which acts by damaging DNA.
XX The sequences within the gene cluster, and the proteins they encode, can
XX
XX Query Match 13.1%; Score 242.4; DB 3; Length 63164;
XX Best Local Similarity 50.3%; Pred. No. 2.7e-37;
XX Matches 749; Conservative 0; Mismatches 706; Indels 33; Gaps 5;
Qy 227 CGTGGCGAGCGCGCGGCGGATTCCTTCCCTCCGCGCGCGCGCGCGTCCGCGTGC 286
Db 40906 CGTCGCGAGGAGACGCGGACGGTGCATCCCGCGCGCGCGCGCGCGCGCGCGCG 40847
Qy 287 GTCGGAAGCGCGGCTCGCGCTGTTCATCCGAGCGCGCGCGCGCGCGCGCGCGCGCG 346
Db 40846 GAGCGGGAGATCTTCGAAGGATCGCGACAGAAATCCCATCTACGGGGTGACCAC 40787
Qy 347 CGGCTTCGGTCCCTTCGAAACCGCCCTGATCTCAGGTGAGAAATGCCGAACTGCAGGC 406
Db 40786 CGGCTACGGCGAGATGATCTACATGCAGGTGCGACAAGTCGAAGGAGTCGAACCTGCAGAC 40727
Qy 407 CGCATGTTCTGCGCGGTCTGTGTGTCGATCGCTCAGGAGCGCTCCGCTGCAGCGAGG 526
Db 40726 CAATCTCGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 40667
Qy 467 CGCATGTTCTGCGCGGTCTGTGTGTCGATCGCTCAGGAGCGCTCCGCTGCAGCGAGG 526
Db 40666 GGCATGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 40607
Qy 527 GACCATCGCTCGCCTGATCGACCTTGCTCAATTCGAGCTCGCTCCGCGCGTCCAGCGCG 586
Db 40606 CATCTTCGAACGCTTCGCGCGAGTACCTGAAACGAGGGCATCACCCCGCGCATACCCGAGAT 40547
Qy 587 CGGACCGGTGGCGCGCTCGGGTGACCTGACACCGCTTCGGCATATGCTGCTCGCTCCA 646
Db 40546 CGGGTCACCTCGGGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 40487
Qy 647 GGGCGGGGAGACTTCTCGGACCGGAGCGGAGCGGCTTGACGGCGCGAGAGGGCTCCG 706
Db 40486 CGGAGGGGTACGTCTCTG---CCGACGAGCGCGCGGTGAGACCGCCCGAGTGTCTGGC 40430
```

Qy		707	GCGGGACGCTCGAACCGTTCGATCTCTCCATCCGATGCAGTGTGGCGCTGGTCAACGG	766
Db		40429	CGACGGGGCATCGAGCCGCTCGAACTGGCGTTCAAGGAGGCCCTCGCATGATCAACGG	40370
Qy		767	GACTCTCGGCATCACCGGATCGCGTGTTGAATGCTCAGCGCTCGCGCACTCTCGGCAA	826
Db		40369	CAGTCCGGGATGACCGGTCTGGGCTCCCTGGTCTCGACGGGCCCTGGAGCAGGCCCA	40310
Qy		827	CTGGGGGTTGGCGTTTACCGGCCCTGTCTTGGGAATGTCTTGAGAGCGCGGACCGAGGCATG	886
Db		40309	GCAGGCCAGATCGTGA CGSCTCTGCTCATCGAGGCGGTACCGCGATCGACCAGCCCCTT	40250
Qy		887	GGCGGGCGACTGTCCGACTCT---GGGCGCGCATCCCGACAGAAGGACCGCGACGAG	943
Db		40249	CTTCGGGGAGGGGACACACATAGCCCCCGCACGAGGGGCCAGATCGACACCGCGCCCAA	40190
Qy		944	GCTCGCGCCCGCGTGGACGGCAGCGCGGGGTGGTCCGGGACGTCATTGCGCAGCGGAG	1003
Db		40189	CATGCGGGCCCTGATCGGGGACGCGA CTGACGCTCGAGCACGCGACCTGCGCCGAGA	40130
Qy		1004	GCTC-----GACGCGCGCGATATCGGGACGGAGCGCGAGGCGGGGACGAGA	1048
Db		40129	ACTCCAGAAAGACAAGAGGCGCGCAAGGACGTC CAGCGCTCGAGATCTACTCTGACAA	40070
Qy		1049	TGCTACAGCCTCGCTCGCTCCGCA GGTTCCTCGGGCGSGCTTCGACAGCTCGCATG	1108
Db		40069	GGCTTACTCGCTCGGGGCATATCCCCAGGTCTGTGGGGCGGTGCGGACACCTTTGACCA	40010
Qy		1109	GCATGACCGGGTCTGACATCGAGCTGAACCGGCTGACCGACAATCTCGGTGTTTTCCGCC	1168
Db		40009	CGCGGSCACAAGCTGCGATCGAGCTCAACTCGGCCAACGACAACCCGCTCTT-----	39956
Qy		1169	CGATGGAGGTTGCCCGCTTCGACGGGGCAATTTATGGGCCAGCATGTGGCGCTGAC	1228
Db		39955	CTTCGAGGGCAAGGAGATCTTTCACGGGGCGAACTTCCA CGGTACGCCGATGCGGTTCGC	39896
Qy		1229	GTCGGATGCTCGCCACAGGCGCTCACCTTCTGGCGGGCTTTGCGGAGCGCCAGATTGC	1288
Db		39895	GATGGACTTCGTGACCATCGCGCTCACCCAGCTTCGGGCTCTTGGCGGAGCGCAGATCAA	39836
Qy		1289	ACGTCTGACAGATGA AAGCTGAACCGTGGGTGCCCCCTTCCTCCACCGGGGCCCGCC	1348
Db		39835	CCGGTCTCTGAACCGGCACCTCAGCTACGCGCTCCCGAGTTCTCTGCTCCGGGACCC	39776
Qy		1349	CGGGTTGAATTCGGGCTTCATGGGCGCACAGGTGACGGCGACCGCGCTCTGGCCGAGAT	1408
Db		39775	GGGGCTGACAGCGGATTTGCGGGCGCCAGTAGTACCGGCCACCGCACCTGGTGGCCGAGAA	39716
Qy		1409	GCGAGCCACGCGGACCTGCGCTCGATTCGATCTCTCACGAACGCGGCCAATCAGGATGT	1468
Db		39715	CCGGACGATCGGCCCGCGCAGCACCCAGAGCGTCCCGTCCAACGGGACACACGAGACGT	39656
Qy		1469	GGTCTCGTTGGGACCATCGCCGCGCGCTCTTGGCGCGAGAAGATCGACCGTTGGGGCGGA	1528
Db		39655	GGTGAGCATGGGCTGATCTCGGCGCGCAACCGCCCGCGGGTCTCTGGAAACAACA	39596
Qy		1529	GATCCTTGCGATCCTCGCTCTGTCTTGACAACTGCGGAGCTGCGCTGCGGACGCGG	1588
Db		39595	GATCCTCGCG-----GTGGAGTACTTGGCGCGGCCACAGGCGGTGCAATCTCGGCCG	39542
Qy		1589	CCTAGACGGGGTGTCTCCCGCGGGGAAGAGCTGTGTG CAGGGCCCTCGCGAGCAGTTCCC	1648
Db		39541	GTTTCGACGGCTTGAGCCCGCGCGGAAGGCCACGTACGAAGCGGTGCGCGGCTGTTCC	39482
Qy		1649	GCGCTTTGACACGACCGGCCCTCTGGGACAGGAAATTGCGCGCTTGC	1696
Db		39481	GAGCTGTGGCGTGCACCGGTATATGTCGACGACATCGAGCTGGTGC	39434

RESULT 7
ADS55996
ID ADS
XX

AD55996;
02-DEC-2004 (first entry)
Bacterial polynucleotide #7983.
Recombinant DNA construct; transformed plant; improved plant property; cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis; pathogen tolerance; pest tolerance; plant disease resistance; cell cycle pathway modification; plant growth regulator; homologous recombination; seed oil yield; protein yield; carbohydrate; nitrogen; phosphorus; photosynthesis; lignin; galactomannan; bacterial polynucleotide; gene; ss.
Bacteria.
US2003233675-A1.
18-DEC-2003.
20-FEB-2003; 2003US-00369493.
21-FEB-2002; 2002US-0360039P.
(CAOY/) CAO Y.
(HINK/) HINKLE G J.
(SLAT/) SLATER S C.
(CHEN/) CHEN X.
(GOLD/) GOLDMAN B S.
Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
WPI; 2004-061375/06.
New recombinant DNA construct comprising a promoter positioned to provide for expression of a polynucleotide encoding a polypeptide from a microbial source, useful for producing plants with improved properties.
Claim 1; SEQ ID NO 31670; 122pp; English.
The invention relates to a recombinant DNA construct comprising a promoter functional in a plant cell, where the promoter is positioned to provide for expression of a polynucleotide encoding a polypeptide from a microbial source. The invention also relates to a transformed plant comprising the recombinant DNA construct and a method of producing a transformed plant having an improved property. The plant is a crop plant such as maize or soybean. The method of producing a transformed plant having an improved property comprises transforming a plant with the recombinant DNA construct and growing the transformed plant, where the polynucleotide or polypeptide is useful for improving plant properties. The recombinant DNA construct is useful for producing plants with improved plant properties, e.g. improved cold, heat or drought tolerance, tolerance to herbicides, extreme osmotic conditions, pathogens or pests, increased resistance to plant disease, better growth rate by modification of the cell cycle pathway with plant growth regulators, increased rate of homologous recombination, modified seed oil or protein yield and/or content, improved yield by modification of carbohydrate, nitrogen or phosphorus use and/or uptake, by modification of photosynthesis or by providing improved plant growth and development under at least one stress condition. improved lignin production or improved galactomannan production. This sequence represents a bacterial polynucleotide used in the scope of the invention. Note: The sequence data for this patent did not form part of the printed specification but was obtained in electronic format from USPTO at seqdata.uspto.gov/sequence.html.
Sequence 1497 BP: 191 A: 577 C: 527 G: 202 T: 0 U: 0 Other:

Query Match 12.9%; Score 238.4; DB 13; Length 1497;
Best Local Similarity 50.8%; Pred. No. 1.4e-36;
Matches 739; Conservative 0; Mismatches 666; Indels 51; Gaps 5

Qy 199 GCACATCATGATCTGGACACAGCCCATGCCGTGGCAGCGCGCGCGGATTCCTTG 258

Db 14 GCAGAGCAGCTCGCGCAGCTCGAAGCGATCTGGCGAGGGGGCGCGCGCTCTGG 73
Qy 259 CCCCTCCGGCGGCGACCGGTGCGTTCGAAGACGCGGCTCGGCGCTGTCTATCCCGC 318
Db 74 CCCCCGAGCGCGCCCGCGCTCGAGAGGGCGCGCGCGCTGGCGAGGCGCGCGCG 133
Qy 319 AGGCGCGCATGTCTACGAGCTGACAAACCGGCTTCGGTCCCTTGGCGAACCGCTGTATCT 378
Db 134 GCACGGCGCGGTCTATGGGTGAACAGGGCTTCGGCAAGCTTCGGAGCTCTAAGATCG 193
Qy 379 CAGGTGAGAAATGTCCGAACGCTGACGGCCAACTTTGTCCATCATCTGGCCAGCGCGGTGG 438
Db 194 CTCGGCGCATACGGCGCAACTGCAACGCAACTGATCTGTGCACTGTCTGGCGGTGG 253
Qy 439 GACCGGTCTTGAATGAGAGCGGCGCGCGCATGGTCTTGGCGGTCTGTGTGATCG 498
Db 254 GCGAGCTATGCCCCCGCTCCACGGCGCGGCTGATGATTCGCTGAAGCTCTGTGCTCG 313
Qy 499 CTCAGGAGAGCTTCGGTCCAGCGAGGGGACCATCGCTCGCTGATCGACCTGTCAATT 558
Db 314 GCGCGGCGCTTCGGCGTGGCTGGAGATGTGGCGCTGCTCGAAGGCAATGTGGCGG 373
Qy 559 CCGAGCTCGCTCCGGCGCTTCCAGCCCGGACAGGTGGGCGCGTTCGGGTGACCTGACAC 618
Db 374 CGGCGTCAACCGCGTATCCGGCGCAGGGTGGTGGCGGCGAGCGCGATCTGGCAC 433
Qy 619 CGCTTGGCATATGGTGTCTGCTTCCAGCGCGGAGAGCTTCTGGACCGGGACGGGA 678
Db 434 CCCTCGCCCATATGGCCCGCTGATGATCGCGAGGGCGAGGCGAGGTC---GGCGCA 490
Qy 679 CGCGCTTACGGCGCAAGAGGCTCCGGCGCGAGCGCTGCAACCGCTCGATCTCTCC 738
Db 491 GCGCGCTCCCGGTGCGCGGCTTGGCGAGCGCGTCTTGGCCCGGTGGCCCTCGGAC 550
Qy 739 ATCGCATGCACTCGCTGTCTCAACGGGACCTCCGCCATGACCGGGATCGCGTGTGA 798
Db 551 CCAAGGAAGGCTCGCCCTCATCAAGGCAACGCAATCTCGACCGCTATGCCCTCGCG 610
Qy 799 ATGCTACCGCTCGCGCATCTCGGCAACTGGCGGTGGCGGTGACGGCCCTGTTCGG 858
Db 611 GCCTCTTCGAGGGTGGCGCGCGCTCAGGGCGCGCTGTGATCTCGGCCCTCTCCACG 670
Qy 859 AATGCTGAGAGCGGACCGAGCATGCGGCGCGGCACTGTCGACCTGGCGCGCATC 918
Db 671 ATGGATATGGGTGCACTCGCGCGCTCCGCCCGAGATCCAGCGCTGGCGCGCATG 730
Qy 919 CCGGACAGAGGCGCGCGAGGCTCGCGCGCGCGCGTGGAGCGGCGCGCGGTGG 978
Db 731 CGGCGCATGCGAGGGCGCGCCACCATGCGCGCGCTGCTCGAAGGCTCGGCCATCGCG 790
Qy 979 TCGGCGAGTCAATTCGAGCGGAGGCTCGACCGCGCGATATCGGAGCGGAGCGGAGG 1038
Db 791 AGAGCCACCGTAGGGGAC-----AGC 814
Qy 1039 CGGCGCAGGATGCTACAGCTCGCTCGCTCGCGAGGTTCTCGGGCGGCTTCGACA 1098
Db 815 GGGTGCAGGACCCCTACTGCACTCGCTCGCAGCGCGAGTGACGGGCGCGCGATGATG 874
Qy 1099 CGCTCGCATGTCATGACCGGTGTGACGATCGAGCTGAACCGCGGTGACCGCAATCCGG 1158
Db 875 TGCTCGCATGCGCGGCGGACGCTGGCCACCGAGGCAATGCCCGCACCGCAATCCG 934
Qy 1159 TGTTTCGCGCATGAGCGAGTGGCGCGCTCGCAGCGGGCAATTTATGGGCGAGATG 1218
Db 935 TCGTGTCTCGA-----CGGCGCATGCTGCTCGGAGGCAACTTCATGCCGAGCGCG 988
Qy 1219 TGGCGCTGACGTTCGATCGCTCGCAGCGGCGGTCACTGCTTGGCGGCTTCGGAGC 1278
Db 989 TGGGCTTCGCGCGGACATGATCGCTGGCGCTCTCGAGATCGGCGCCATCGCGAGC 1048
Qy 1279 GCCAGATTGACGCTGTGACACATGAAGGCTGAACCGGTGGGCTGCCCGCTTCTCCACC 1338
Db 1049 CCGCGTGGCGCTATGTTGATCGGACGCTCTCTTCGACCTTCGCGCTTCTTCACCC 1108

RESULT 8

ACA25972

ID ACA25972 standard; DNA; 1521 BP.

XX ACA25972;

XX AC ACA25972;

XX DT 19-JUN-2003 (first entry)

XX DE Prokaryotic essential gene #7629.

XX KW Antisense; ds; prokaryotic essential gene; cell proliferation;

XX KW drug design; gene.

XX OS Burkholderia mallei.

XX PN W0200277183-A2.

XX PD 03-OCT-2002.

XX PF 21-MAR-2002; 2002WO-US009107.

XX PR 21-MAR-2001; 2001US-00815242.

XX PR 06-SEP-2001; 2001US-00948993.

XX PR 23-OCT-2001; 2001US-0342923P.

XX PR 08-FEB-2002; 2002US-00072851.

XX PR 06-MAR-2002; 2002US-0362699P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

XX PI Wall D, Trarwick JD, Carr GU, Yamamoto R, Forsyth RA, Xu HH;

XX WPI; 2003-029926/02.

XX DR P-PSDB; ABU22102.

XX PT New antisense nucleic acids, useful for identifying proteins or screening

XX PT for homologous nucleic acids required for cellular proliferation to

XX PS isolate candidate molecules for rational drug discovery programs.

XX PS Claim 14; SEQ ID NO 13842; 1766pp; English.

XX CC The invention relates to an isolated nucleic acid comprising any one of

XX CC the 6213 antisense sequences given in the specification where expression

XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:

XX CC (1) a vector comprising a promoter operably linked to the nucleic acid

XX CC encoding a polypeptide whose expression is inhibited by the antisense

XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated

CC polypeptide or its fragment whose expression is inhibited by the
CC antisenese nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisenese nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 1521 BP; 220 A; 548 G; 530 G; 223 T; 0 U; 0 Other;

Query Match 12.5%; Score 231.6; DB 8; Length 1521;
Best Local Similarity 50.9%; Pred. No. 2.8e-35;
Matches 753; Conservative 0; Mismatches 664; Indels 63; Gaps 6;

QY 229 TGGCGAGCGCGCGCGGATGCTCTCCCTCCGCGCGCGCGCGCGTGGCGT 288
DB 50 TCGCCCCGAGAACGTGCAGATCGCTCGATCCCGGAGCTTCGCGCGATGACCGGG 109
QY 289 CCGAAGCGCGGCTCGGCGCTGTATCCGCGAGGCGGCCATGTCTACGAGCTGACAAACCG 348
DB 110 GCGCGAGGCGCTCGCGACATCGCCGGAAGGCGGCGGTACGGCATCAACACGG 169
QY 349 GCTTCGGTCCGCTTGGGAAACCGCTGTATCTCAGGTGAGATGTCCGAAACGCTGAGGCCA 408
DB 170 GCTTCGGGCGCTCGCGAGCAGCACATTCGACACGACCGCTCGAGCTCTGCAGAGA 229
QY 409 ATCTTGTTCATCATCTGCGCAGCGGCTGGGACCGGCTTGTGACTGACACGCGCGCG 468
DB 230 ACTTCGTGCTGTGCGACCGGCTGGGCTCGCGAGCGCGATGGCGCGCCCGTGTGCGCC 289
QY 469 CCATGGTTCCTGGCGCTCTGGTGTGCTATCGCTCAGGAGCGCTCGGCTGCCAGCGAGGGA 528
DB 290 TGTGTATGGCGCTCAAGCTCTCGAGCTTCGCGCGGCGCACTCGGGCATTCGTGCGGTG 349
QY 529 CCATTCGCTCGCTGATCGACTGCTCAATTCGAGCTCGCTCGCGCGGCTTCCAGCGCGG 588
DB 350 TGATGGACGCGCTCGTCAACGCTGTTCAACGCGGACGCTGCTCGCGCTCATTCGCGTCAA 409
QY 589 GCACGCTGGGCGGCTCGGCTGACTGACACGCTTGGCGATATGGTCTCTGCTCCAGG 648
DB 410 GCTCGGTGGGCGGCTCGGCGGCACTTCGCGCGCTCGCGCGCATGTGCGCGCTGTGCTG 469
QY 649 GCGCGGAGACTTCCTGGACCGGACCGGACCGGCTTGTGACCGCGCAGAGGCGTCCGGC 708
DB 470 GCATCGGCGAGCTGTTCATTCGGGCGCA---GCGCGGAGCGGCGCGAAGGAGCTGCGTG 526
QY 709 GCGGACGCGCTGCAACCGCTCGATCTCTCCCATCGCGATGCACTGGCGCTGCTCAACGGGA 768
DB 527 TCGCGGCGCTCGCGCGCTCAACGCTCGAAGGAGGAGGCGCTCGCGCTGCTGAAACGGCA 586
QY 769 CCTTCGCGCATGACCGGATCGCTGTGTAATGCTACGCGCTGCGCGCATCTCGGCACT 828
DB 587 CGCAGGCGTGCAGCGCGCTCGCGCTGCACAACTGTTTCGCGATCGAGGACCTGTACCGCA 646
QY 829 GGGCGGTGGGCTTGACGCGGCTGCTTCGCGAATGCTCTGAGAGGCGCGACCGAGGCATGGG 888

RESULT 9
ADT44460

ID ADT44460 standard; cDNA; 1503 BP.

XX AC ADT44460;

XX DT 02-DEC-2004 (first entry)

XX DE Bacterial polynucleotide #19211.

XX KW Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW pathogen tolerance; pest tolerance; plant disease resistance;

cell cycle pathway modification; plant growth regulator; homologous recombination; seed oil yield; protein yield; carbohydrate; nitrogen; phosphorus; photosynthesis; lignin; galactomannan; bacterial polynucleotide; gene; ss.

Bacteria.

US2003233675-A1.

18-DEC-2003.

20-FEB-2003; 2003US-00369493.

21-FEB-2002; 2002US-0360039P.

(CAOY/) CAO Y.

(HINK/) HINKLE G J.

(SLAT/) SLATER S C.

(CHEN/) CHEN X.

(GOLD/) GOLDMAN B S.

Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS; WPI; 2004-061375/06.

New recombinant DNA construct comprising a promoter positioned to provide for expression of a polynucleotide encoding a polypeptide from a microbial source, useful for producing plants with improved properties.

Claim 1; SEQ ID NO 42898; 122pp; English.

The invention relates to a recombinant DNA construct comprising a promoter functional in a plant cell, where the promoter is positioned to provide for expression of a polynucleotide encoding a polypeptide from a microbial source. The invention also relates to a transformed plant comprising the recombinant DNA construct and a method of producing a transformed plant having an improved property. The plant is a crop plant such as maize or soybean. The method of producing a transformed plant having an improved property comprises transforming a plant with the recombinant DNA construct and growing the transformed plant, where the polynucleotide or polypeptide is useful for improving plant properties. The recombinant DNA construct is useful for producing plants with improved plant properties, e.g. improved cold, heat or drought tolerance, tolerance to herbicides, extreme osmotic conditions, pathogens or pests, increased resistance to plant disease, better growth rate by modification of the cell cycle pathway with plant growth regulators, increased rate of homologous recombination, modified seed oil or protein yield and/or content, improved yield by modification of carbohydrate, nitrogen or phosphorus use and/or uptake, by modification of photosynthesis or by providing improved plant growth and development under at least one stress condition, improved lignin production or improved galactomannan production. This sequence represents a bacterial polynucleotide used in the scope of the invention. Note: The sequence data for this patent did not form part of the printed specification but was obtained in electronic format from USPTO at seqdata.uspto.gov/sequence.html.

Sequence 1503 BP; 225 A; 535 C; 516 G; 227 T; 0 U; 0 Other;

Query Match 11.6%; Score 215.4; DB 13; Length 1503;

Best Local Similarity 50.0%; Pred. No. 3.7e-32;

Matches 753; Conservative 0; Mismatches 691; Indels 63; Gaps 6;

203 CATCGATCTGACACGAGCCCATGCGTGGCGAGCGGGCGCGGGAGTGTCTTGGCCCC 262

21 CCTGAAGCTCGAGGAATCTCCAGGTGGCTCGCAACGAGGCCACCGTGGAGCTGTGCC 80

263 TCCGGCGGCGACCGGTGCCGTGCGTCCGAAGCGGGGTTCGGCGCTGTTCATCCGCGAGGC 322

81 CGAGCGCGCCACCGCGGTGCGGGGCTCGCGCGGCTGTGGTACCGCGGTTCGCGCGGAGA 140

323 GCGGCATCTTACGGAATGCAACCGGCTTCGGTCCCTTCGGAACCGCTGTATCTCAGG 382

141 CACGCCCGCCCTACGGCATCAACACCGGCTTTGGCACGCTTGGCCGAGGTTCGCGCATCGCAA 200

Qy 383 TGAGATGTCCGAACGCTGACGGCCAAATCTTGTTCATCATCTGGCCACGCGGCTGGGACC 442

Db 201 GAAGGACCTGCGCGACCTCCAGCGCAACCTATCTCTCCACACGCTGTGGCGTGGGCAC 260

Qy 443 GGTGCTTACTGAGACGACGGCGCGCCATGTTCTTGGCGGCTCTGTGTGTCATCGCTCA 502

Db 261 GCCCTTCCCTTCCGGAAGCGCGGGGCTCTCTGTGCTGCTGCAACGTGTCTCGCCAA 320

Qy 503 GGGAGCTTCGGTTCAGAGCGAGGGAGCCATCGCTCGCTGATGACACTGTCTCAATTCGA 562

Db 321 GGGCTACTTCGGGCATCCGCATGAGAGCGTGGCCCTTGGCGCTGGACATGCTGNAACCGGA 380

Qy 563 GCTCGCTCCGGCCGTTCCAGCCCGGGGACAGTGGGCGCGCTGGGTGACCTGACACCGCT 622

Db 381 CGTGTGCGCGGTGCTCCCGAGCGGGGACGCTGGCGGCTCCGGGATCTCGCCCGCT 440

Qy 623 TGGCATATGCTGCTCTCGCTCCAGGCGCGGGAGACTTCTGTGACCGGGAGCGGACCG 682

Db 441 GCGGCACTGCGCTCTGTTTCATCGCGGAAGGTGAAGCTTCTATCAGGGCCA---GCG 497

Qy 683 GCTTGACGGGCGAGAGGGCTTCGGCGCGGAGCGCTGCAACCGCTCGATCTCTCCCATCG 742

Db 498 GATGCCCGGAAGCAGGCGCTGGAGCGCGCGGCTGCAACCGGTGTGTGGAGCCAA 557

Qy 743 CGATGCACTGGCGTGTGTAACGGGACCTTCGCCCATGACCGGATCGCGCTGTGTGATGC 802

Db 558 GGAGGGCTCGCGCTGTGTAACGGGACACAGGCCATGTGCGCGTGTGGCACCTCTCTCCA 617

Qy 803 TCACGCTGCGCGCCATCTCGGCACTGGGGGTGCGCTGTGACGCCCTCTTGGCGGATG 862

Db 618 GTTTCGCGGGAGTTCCTTGGCGGACATCGCGCGCGCCATGACGCTGGAGGG 677

Qy 863 TCTGAGAGCGCGGACCGAGGCTGGCGCGCGGCACTGTCCGACCTGGCGCGCATCTCCGG 922

Db 678 GCTGCTGGGAAGCCACAAGCCCTTCTTCTGAGATTACAGACCTCCGCGCGCACCGGG 737

Qy 923 ACAGAGGACCGCGACGAGGCTTGGCGCGCGCGGTGGAACGCGAGCGCGCGGTGTGTCG 982

Db 738 CCAGAGGAGCTTCGCGCGGACCTTCGCGCGCATCTCTGTGTGACAGCGAGCTGTGGAGTC 797

Qy 983 GCAGCTATTGCGGAGCGGAGGCTCGAGCGCGCGGATATCGGACGAGCGCGGAGGGGG 1042

Db 798 GCAGCTCA-----ACTCAGCAAGGT 818

Qy 1043 GCAGGATGCTTACAGCTGCGCTGCGCTCGCGAGGTTCTCGGGCGCGGCTTCGACACGCT 1102

Db 819 GCAGGACCTTACTCTCTGCGCTGATGCGCGAGGTGCACGGCGCGCGCGCGGAGGCAT 878

Qy 1103 CGCATGGCATGACCGGGTGTGACGATCGAGCTGAACGCGGTGACGACAAATCCGGTGT 1162

Db 879 CGCGTCTCCCGGCGCATCTCGAGGTGGAGGTCAACAGCGCGGACGACAAACCGGCTCGT 938

Qy 1163 TCCGCCGATGGGAGCTGCGCGCGCTGACGCGGGGCAATTCATGGCGCAGCATGTGGC 1222

Db 939 ---GTTCCGAGACACGAGGCGCATCTGTCTGGGGCGCAACTTCACGCGCGCCCATCTC 995

Qy 1223 GCTGACCTCCGATCGCGCTCGCCACGCGCGCTCACCGTTCTGGCGGGCTTGGGAGCGCA 1282

Db 996 CTTGGCCATGACGCTGTGGGATGGCGCTGACGCAACTGTCTGTCATCAGCAGAGCGCG 1055

Qy 1283 GATTGCACTGTGACAGATGAAGGCTGAACCGGTGGGCTGCCCCCTTCTTCTCCACCGGG 1342

Db 1056 CGTGGAGCAGCTGTGAACCCGCTGCTGTCAACCTGCGCGCT-----TCCTGGCGAA 1109

Qy 1343 CCGCGCGGCTTGAATTCGGGCTTCATGGGGGACAGGTGACGCGGACCGCGCTCTGGC 1402

Db 1110 GAATCCGGGTTGAATCTCCGCTTATGATGCGCGAGGTGACCAAGCGCGCTGGTGGC 1169

Qy 1403 CGAGATCGGAGCCACGGGA---CCTGCTCGATCCATTTCGATCTCCAGAACCGCGCCAA 1459

Db 1170 CGAGTCCCGCTGTGTGAGCCACCCCGCTCGGTGGATTGATTCGTCATCCGCGGGCG 1229

QY 1460 TCAGGATGTGGTCTCGCTTGGGACCATCGCGCGCGCTCTGCGCGAGAGATCGACCG 1519
DB 1230 AGAGACACCGTGTCCATGGGCATGACGCGCGCTCAAGGGCGGTGAGGTTCAGCGACTT 1289
QY 1520 TTGGGGGAGATCCTTGGATCCTCGCTCTCTGTCTTGGACAGCTCGGAGCTGGCGTG 1579
DB 1290 CGCCCGTTCGTGCTCGCGATTGAAATTCCTGTGTGGCGCGAGGCCCTGAGACTTCCGCC 1349
QY 1580 CGGACGGGCTAGACGGGGTGTCTCCCGGGGGAAGAAGCTGTGTGAGGCGCTTGCACGA 1639
DB 1350 GCCGCTGAGCCCGCAAGGGCGCTTGGCGGTACGAGCTGG-----TGGCGTC 1400
QY 1640 GCATTTCCCGCGCTTGGAGCGACCGCGCCCTTGGGACAGGAAATTCGCCGCGCTTGTATC 1699
DB 1401 GAAGGTCCCCACATGACAGGAGCGGAGCTGCACCGGACATCGAGGCGGTGAGCCA 1460
QY 1700 GCACCTC 1706
DB 1461 GTCGTC 1467

RESULT 10
ACL71540
ID ACL71540 standard; DNA; 1527 BP.
XX
XX ACL71540;
XX
XX 02-JUN-2005 (first entry)
XX
XX M. xanthus gene sequence, seq id 8003.
XX
XX Transgenic plant; DNA replication; gene regulation; gene expression;
XX gene; ds.
XX
XX Myxococcus xanthus.
XX
XX US6833447-B1.
XX
XX 21-DEC-2004.
XX
XX 10-JUL-2001; 2001US-00902540.
XX
XX 10-JUL-2000; 2000US-0217883P.
XX
XX (MONS) MONSANTO TECHNOLOGY LLC.
XX
XX Goldman BS, Hinkle GJ, Slater SC, Wiegand RC;
XX
XX WPI; 2005-028716/03.
XX
XX New substantially purified Myxococcus xanthus nucleic acid molecule
XX encoding a nitrite reductase, useful for determining gene expression,
XX identifying mutations in a gene of interest, and for constructing
XX mutations in a gene of interest.
XX
XX Example 2; SEQ ID NO 8003; 25pp; English.
XX
XX The invention relates to a substantially purified nucleic acid molecule
XX encoding a nitrite reductase of SEQ ID NO. 11926. Further disclosed is a
XX recombinant DNA construct for expression of a nitrite reductase gene in a
XX plant cell, and a plant cell comprising the recombinant DNA construct.
XX The nucleic acid is useful for determining gene expression, identifying
XX mutations in a gene of interest, and for constructing mutations in a gene
XX of interest. Sequences given in records for SEQ IDs 1850-9691 represent a
XX set of about 7842 genes or partial genes from the genome of the bacterium
XX Myxococcus xanthus. Note: The sequence data for this patent did not form
XX part of the printed specification, but was obtained in electronic format
XX directly from USPTO

SQ Sequence 1527 BP; 228 A; 545 C; 523 G; 231 T; 0 U; 0 Other;
Query Match 11.6%; Score 215.4; DB 14; Length 1527;
Best Local Similarity 50.0%; Pred. No. 3.7e-32;

Matches 753; Conservative 0; Mismatches 691; Indels 63; Gaps 6;
QY 203 CATCGATCTGGACACAGGCCCATCGCGTGGCGAGCGCGCGGATTCCTTGGCCCC 262
DB 36 CCTGAAGCTCGAGGAATCTCCAGGTGGTCCGACGAGGCCACCGTGAAGTGTGCC 95
QY 263 TCCGGCGCGACCGGTGCGTCCGAAAGCGCGCTCGGCGCTGTCAATCCGAGGC 322
DB 96 CGACGCGGCACCGCGTGGCGCTCGCGCCCTGGTGGACCGGTGCGCCCGGAGA 155
QY 323 GCGCATGTCTACGACTGACACCGGCTTCGGTCCCTTGGAAACCGCTGATCTCAGG 382
DB 156 CAGCCCGCGCTACGGCATCAACACCGGCTTGGCACGTGGCGGAGTCCGCAACAA 215
QY 383 TGAAGATGTCGGAACGCTGAGCGCAATCTGTTCATCATCTGCGCAGCGCGTGGGACC 442
DB 216 GAAGGACCTGCGGACCTCGAGCGAACCTTCATCTCTCCACGCGTGTGGCTGGCAC 275
QY 443 GGTGCTTGAATGACGACGCGCGCCATGTGTCTGGCGCGTCTGGTGTGATTCGCTCA 502
DB 276 GCCCTTCCCTTCGGAAGCGCGCGCTCTCTGCTGCTCGCTGCAACGCTGCTGCCAA 335
QY 503 GCGAGCTCCGCTGCGAGGAGGACCATCGCTCGCTGATCGACTGCTCAATTCGA 562
DB 336 GGGCTACTCGGCATCCGATCGGAGCGCTGGCGCTTGGCGCTGGAATGCTGAACCGGGA 395
QY 563 GCTCGCTCCGCGCGTTCCAGCGCGCGCGTGGCGTGGGCTGACCTGACACCGCT 622
DB 396 GCTGTCGCGTGTCCCGAGCGGGGACGCTGGCGGCTCCGGGATCTCCCGCGCT 455
QY 623 TCGCATATGTGTCTCTGCTCCAGGCGCGGAGACTTCTTGGACCGGACGCGGCG 682
DB 456 GCGCACCTGGGCTGCTTTCATCGCGAAGGTGAAGCTTCTATCAGGGCCA---GCG 512
QY 683 GCTTGAAGCGCGAGAGGCTCCGCGCGGACCGCTGCGACCGCTCGATCTTCCCATCG 742
DB 513 GATGCCCGGAAGCAGCGCTGGAGCGCGCGCGCTGCAACCGGTGTGTGGAGGCCAA 572
QY 743 CGATGCACTGGCGCTGTCAGCGGGACCTCCGCGATGACGGGATCGCGTGTGATGC 802
DB 573 GGAGGCGCTCGCTTGTGAACCGGACACAGGCAATGTGCGCGTGGGACACCTGTCTCA 632
QY 803 TCACGCGCTCGCGCATCTCGGCAACTGGCGGTGGGTGTGACGGCGCTGTCTCGGAATG 862
DB 633 GCTTCGCGCGAGTCCCTGCGGACATCGCGGAGCTCGCGGGCGCATGACGCTGAGGG 692
QY 863 TGTGAGAGCGGACCGAGGCAATGGCGCGGCACTGTCCGACCTGCGCGCGCATCCCGG 922
DB 693 GCTGCTGGGAAGCCACAAGCCCTTCAATCTGAGATTCAACGATTCACGATCCGCGCACCCGGG 752
QY 923 ACAGAGGACCGCGAGCGGTGCGCGCCCGTGGACGGCAGCGCGCGGTGTGTCG 982
DB 753 CCAGAGGACGTCGCGCGGACCTCGCGGCGATCTGCGGCGATCTGCGTGGACGAGCTGTGGAGTC 812
QY 983 GCACGTCAATTCGCGAGCGGAGGCTCGACGCGCGCGATATCGGGACGAGCGCGGCGGG 1042
DB 813 GCACGTCA-----ACTGACGAAGGT 833
QY 1043 GAGGATGCTCAGCGCTGCGCTCGCGAGTTCGCGGGCGGGTTCGACAGCT 1102
DB 834 GCAGGACCCCTACTCTCGCTGCAATGCGGAGGTGACGCGCGCGCGCGGAGGCGAT 893
QY 1103 CGCATGCGATGACCGGGTCTGACGATCGAGCTGAAACGCGGTGACCGACAATCCGGTGT 1162
DB 894 CGGTTCTCCGCGCGATCTTGAGGTGGAGGTCAACAGGCGACGAGCAACCCGCTCGT 953
QY 1163 TCCGCCCGATGGCAGCGTGCCTCGCCCTGCAACGCGGGCAATTTTCATGGGCGAGATGTGGC 1222
DB 954 ---GTTGCGGACACGAGGCGCATCTGTGCGGGCGCACTTCCACGCGCAGCCCATCTC 1010
QY 1223 GCTGAGCTCCGATGCGCTCGCACGCGCGCTCACCGTTCGCGGGGCTTGGCGAGGCCCA 1282
DB 1011 CTTGGCCATGGAGCTGTGGCGATGCGGCTGACGCAACTGTCTCATCAGCGCGCGG 1070

Db 2153 GCAGGACCCCTACTCCTCGCTGCATGCGCAGGTGCACGGCGCGCGCGAGGGCAT 2212
QY 1103 CGCATGCGATGACCGGGTGTGACGATCGAGCTGAACCGCGGTGACGACAAATCCGGTGT 1162
Db 2213 CGGTTCTCCGGGCGATCTCGAGGTGAGGTCAACAGCGCGACGACAAACCGCTCGT 2272
QY 1163 TCCGCCCGATGGCAGCGGTGCCCGCCCTGCAACGGGGCAATTTTCATGGCCAGCATGTGC 1222
Db 2273 ---GTTCCGGGACACGAGCGCATGTGTGCGGCGGCAACTTCCAGCGCCAGCCCATCTC 2329
QY 1223 GCTGAGCTCGATGCGCTGCCAGCGCGCTGACCGTTCCTGGCGGGCTTCGCGAGGCCA 1282
Db 2330 CCTGGCCATGAGACGTGTGGCGATGCGCTGACGCAACTGTCTGCTCATCAGCAGCGCG 2389
QY 1283 GATTGACGCTGCACAGATGAAGGCTGAACCGTGGGCTGCGCCCTTCCTCCACCGGG 1342
Db 2390 CGTGGAGCAGCTGTGNAACCGCTGCTGTCAACCTGCGGGCT-----TCTGGCGAA 2443
QY 1343 CCCCGCGGGTTGAATTCGGGCTTCATGGGCGCACAGGTGACGGCGACCGCGCTCCTGGC 1402
Db 2444 GAACTCGGGTTGAATTCGGGCTTCATGATCGCGAGGTGACCGCGCGCTGGTGGC 2503
QY 1403 CGAGATGCGAGCAGCGGA---CCTGCTCGATTCATTCGATCTCCAGCAACCGCGCAA 1459
Db 2504 CGAGTCCCGGTCTGAGCCACCGCGTCCGTGATTCGATTCCTGTCATCCCGCGGCG 2563
QY 1460 TCAGGATGTGTCTGCTGGGACCATCGCGCGCGCTCTGCGCGAGAGATCGACCG 1519
Db 2564 AGAGGACACGTGTCAATGGGCGATGACGGCGCGCTCAAGGGCGCTGAGTCAAGCACTT 2623
QY 1520 TTGGGCGGAGATCCTTGGCATCTCGCTCTCTCTTGCACAAAGCTGCGGAGCTGGCTG 1579
Db 2624 CGCCGTTCTGCTCGGATGAATCTGTGGCGCGCGCTCAAGGGCGCTGAGTTCGCGCT 2683
QY 1580 CGCAGCGGCTAGACGGGTGTCTCCGCGGGAAGAAGCTGGTGAGGCCCTGCGCGA 1639
Db 2684 GCGCTGAAGCCCGGAAGGCGCCCTCGCGGCTACGAGCTG-----TGCGCTC 2734
QY 1640 GCAGTTCCCGCCCTTGAGACGACCGCGCCCTGCGGACAGGAATTCGCGCTTGTCTAC 1699
Db 2735 GAAGGTCCCCCATGACGAAGGACCGGAGCTGACCGGGACATCAGGCGGTGAGCCA 2794
QY 1700 GCACCTC 1706
Db 2795 GCTCGTC 2801

RESULT 12
ADS50013
ID ADS50013 standard; cDNA; 1503 BP.
XX
AC ADS50013;
XX

02-DEC-2004 (first entry)

Bacterial polynucleotide #4756.

Recombinant DNA construct; transformed plant; improved plant property;
cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
pathogen tolerance; pest tolerance; plant disease resistance;
cell cycle pathway modification; plant growth regulator;
homologous recombination; seed oil yield; protein yield; carbohydrate;
nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
bacterial polynucleotide; gene; ss.

Bacteria.

US2003233675-A1.

18-DEC-2003.

20-FEB-2003; 2003US-00369493.

XX

PR 21-FEB-2002; 2002US-0360039P.
XX
PA (CAOV/) CAO Y.
PA (HINK/) HINKLE G J.
PA (SLAT/) SLATER S C.
PA (CHEN/) CHEN X.
PA (GOLD/) GOLDMAN B S.
XX
PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX WPI; 2004-061375/06.
XX
PT New recombinant DNA construct comprising a promoter positioned to provide
PT for expression of a polynucleotide encoding a polypeptide from a
PT microbial source, useful for producing plants with improved properties.
XX
XX Claim 1; SEQ ID NO 28443; 122pp; English.
XX
CC The invention relates to a recombinant DNA construct comprising a
CC promoter functional in a plant cell, where the promoter is positioned to
CC provide for expression of a polynucleotide encoding a polypeptide from a
CC microbial source. The invention also relates to a transformed plant
CC comprising the recombinant DNA construct and a method of producing a
CC transformed plant having an improved property. The plant is a crop plant
CC such as maize or soybean. The method of producing a transformed plant
CC having an improved property comprises transforming a plant with the
CC recombinant DNA construct and growing the transformed plant, where the
CC polynucleotide or polypeptide is useful for improving plant properties.
CC The recombinant DNA construct is useful for producing plants with
CC improved plant properties, e.g. improved cold, heat or drought tolerance,
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC increased resistance to plant disease, better growth rate by modification
CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or
CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition, improved lignin production or improved galactomannan
CC production. This sequence represents a bacterial polynucleotide used in
CC the scope of the invention. Note: The sequence data for this patent did
CC not form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 1503 BP; 261 A; 514 C; 476 G; 252 T; 0 U; 0 Other;

Query Match 11.4%; Score 211.6; DB 13; Length 1503;
Best Local Similarity 51.3%; Pred. No. 2e-31;
Matches 657; Conservative 0; Mismatches 569; Indels 54; Gaps 5;

QY 301 TCGGGCGCTCATCGCGAGCGCGCCATGTCTACGGACTGACAAACGGCTTCGGTCCCC 360
Db 125 TCGCGATATCGCGGAAGGGGAGCGCGCTTACGGCATCAACAGGGCTTTGGGGCC 184
QY 361 TTGCGAACCGCTGATCTCAGGTGAGAAATGTCCGAACGCTGCAGGCCAATTTGTCCATC 420
Db 185 TCGCCAGCAGCATATTCGCGACGACCAACTCGAACTGTCAGCGCAATCTGTGTCT 244
QY 421 ATCTGGCCAGCGGCTGGGACCGGTGCTTACGAGACGCGCGCGCATGTTCTGG 480
Db 245 CTCACGCGTTCGGCGTGGGTGAGCCGATGTCGGTCCGGTCTGCTGATCGCGC 304
QY 481 CGCGTCTGCTGTCGATCGCTCAGGAGCGCTCGCGTCCAGCGAGGGGACCATCGCTCGCC 540
Db 305 TGAACTCTCGAGCCTCGCGCGCCCATTCGGGCATTCGCGTGAAGTATGACGCGC 364
QY 541 TGATCGACCTGCTCAATTCGAGCTCGCTCCGCGCGCTTCCAGCCCGCGACCGTGGCG 600
Db 365 TGATCAGCTGTACACGCGGAGCTGCTCGCGGTGATTCCGGTCAAGGGTTTCGGTGGTG 424
QY 601 CGTGGGTGACCTGACACCGCTTGGCATATGGTGTCTGCGCTCCAGGGCGCGGAGACT 660
Db 425 CATCGGGCGACCTCGCGCGCTCGCGCATATGTCGGCGCGCTGCTCGCGCTCGCGCA-- 482

```
QY 661 TCCTGACCGGACGGGACGGGCTTGACGGCGCAGAGGGCTCCGGCGCGGACGGCTGC 720
D 483 -AGTGTTCGAGAGGGCGAGCGCATGCGCGCACCGAGGGTCTCGCGTCTGCGGCTCA 541
QY 721 AACCGCTCGATCTCTCCCATCGCGATGACCTGGCGCTGCTCAACAGGGACCTCCGCGCATGA 780
D 542 AGCGGCTCAGCTGCAAGCGCAAGAGAGGGGCTGGGCTCTCTGAACGGCAGCAGGCTTGA 601
QY 781 CCGGATCGCGCTGGTGAATGCTCAGCGCTCGCGCATCTCGGCAATGCGGCGGTGGCGT 840
D 602 CGGCGCTCGCGCTGTACACATGTTTCGCCATCGAAGACCTGTACCGCACCGCGCTGGTGT 661
QY 841 TGACGGCGCTCTTCGCGAATGTTCTGAGAGCGCGGACCGAGCGCATGGGCGCGGCACTGT 900
D 662 CGGCGCATGTTGCTAGTAGTCCGCAATGGGCTGCTCAAGCCGTTCAACGGCGGCAATTC 721
QY 901 CCGACCTCGCGCGCATCCCGGACAGAGAGCGCGCAGCGAGGCTGCGCGCGCGCGTGG 960
D 722 ACGAGTTGCGCGGCGCATCAAGGTCAGATCGACGCGGCGGCGGCATACCGCTCGCTGCTGG 781
QY 961 ACGGACGGCGCGGTGGTTCGGGACGTCATTTGCGAGCGGAGGCTCGACCGCGCGATA 1020
D 782 AAGGTTGCGCAATCAACGTTTTCGCACG----- 808
QY 1021 TCGGAGCGGAGCGGAGCGGCGGAGGATGCTACAGCTCGGCTGCGCTCGCGAGTTC 1080
D 809 -----CGATTGCGACAGGTCGAGGACCCGTACAGCTCGGCTGCGAGCGCGAAGTCA 862
QY 1081 TCGGCGGCGGCTTCGACACGCTCGCATGGCATGACCGGCTGCTGACATCGAGCTGAACG 1140
D 863 TGGGCGGCTGCTGATCAGATCGCCACGCGGCGCAAGTCTGCTCTGACAGCGAAGC 922
QY 1141 CGGTGACGCAATCCGCTGTTTCGCGCGATGAGCGGTCGCCCGCTGCAACGCGGCGCA 1200
D 923 CGGTCTCGCAATCCGCTGATTTTCCCGCA---CACCGGCGAAGTCTGTCGCGGCGGTA 979
QY 1201 ATTTCATGGGCGCAGATGTGGCGTGCAGTCCGATGCGCTGCGCACGGCGCTACCGTTC 1260
D 980 ACTTCCACGACGACCGGTCGCTGCGCGCGGATACCTCGCGCTCGCGCGCGCGGAA 1039
QY 1261 TGGCGGCGCTTGGGAGCGCGCAGATTCACGCTGACAGATGAAGGCTGAACCGTGGCG 1320
D 1040 TCGGCGGCTGCGGACGCGCATCGGCTGCTGATGCGACGCGCTGTGCGGCGCTGC 1099
QY 1321 TGCCCCCTTCTCTCACCGGCGCGCGCGGTTGATTCGGCTTCATGGCGCGCACAGG 1380
D 1100 CGCC-----GTTCTCTGTCGCGATGCGCGCGTGAATTCGGGCTTCATGATCGCGCAG 1153
QY 1381 TGACGCGACCGCGCT---CCTGCGCGAGATGCGAGCGACGGACCTGCTCGATTCATT 1437
D 1154 TCAGGCTGCGCGCTCGCTTCGGAACCAAGACGCTCGCGCATTCGGGCTTCGGTCTGATT 1213
QY 1438 CGATCTCCAGAACCGCGCAATCAGATGTGCTGCTTGGGACCATCGCGCGCGCC 1497
D 1214 CGTGGCCACTTGGCAACACAGAGAACCAAGTGTGATGGCGACGTTTGGCGCGCGCA 1273
QY 1498 TCTGCGCGGAGAGATGACCGCTTGGCGGAGATCTTGGCGATCTTGGCTCTCTGCTTG 1557
D 1274 AGCTCGCGGACATGCTGAAACACCGCGGAACATCTCTGATGATGTAAGTCTGCGCGCG 1333
QY 1558 CACAGCTCGGAGCTGGC 1577
D 1334 CGCAAGGCGTCAATCTGCGC 1353
```

RESULT 13

AD55528

ID AD55528 standard; cDNA; 1503 BP.

XX AC

AD55528;

XX AC

DT 02-DEC-2004 (first entry)

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

Bacterial polynucleotide #7515.

Recombinant DNA construct; transformed plant; improved plant property; cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis; pathogen tolerance; pest tolerance; plant disease resistance; cell cycle pathway modification; plant growth regulator; homologous recombination; seed oil yield; protein yield; carbohydrate; nitrogen; phosphorus; photosynthesis; lignin; galactomannan; bacterial polynucleotide; gene; ss.

Bacteria.

US2003233675-A1.

18-DEC-2003.

20-FEB-2003; 2003US-00369493.

21-FEB-2002; 2002US-0360039P.

(CAOY/) CAO Y.

(HINK/) HINKLE G J.

(SLAT/) SLATER S C.

(CHEN/) CHEN X.

(GOLD/) GOLDMAN B S.

Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;

WPI; 2004-061375/06.

New recombinant DNA construct comprising a promoter positioned to provide for expression of a polynucleotide encoding a polypeptide from a microbial source, useful for producing plants with improved properties.

Claim 1; SEQ ID NO 31202; 122pp; English.

The invention relates to a recombinant DNA construct comprising a promoter functional in a plant cell, where the promoter is positioned to provide for expression of a polynucleotide encoding a polypeptide from a microbial source. The invention also relates to a transformed plant comprising the recombinant DNA construct and a method of producing a transformed plant having an improved property. The plant is a crop plant such as maize or soybean. The method of producing a transformed plant having an improved property comprises transforming a plant with the recombinant DNA construct and growing the transformed plant, where the polynucleotide or polypeptide is useful for improving plant properties. The recombinant DNA construct is useful for producing plants with improved plant properties, e.g. improved cold, heat or drought tolerance, tolerance to herbicides, extreme osmotic conditions, pathogens or pests, increased resistance to plant disease, better growth rate by modification of the cell cycle pathway with plant growth regulators, increased rate of homologous recombination, modified seed oil or protein yield and/or content, improved yield by modification of carbohydrate, nitrogen or phosphorus use and/or uptake, by modification of photosynthesis or by providing improved plant growth and development under at least one stress condition, improved lignin production or improved galactomannan production. This sequence represents a bacterial polynucleotide used in the scope of the invention. Note: The sequence data for this patent did not form part of the printed specification but was obtained in electronic format from USPTO at seqdata.uspto.gov/sequence.html.

Sequence 1503 BP; 261 A; 514 C; 476 G; 252 T; 0 U; 0 Other;

Query Match 11.4%; Score 211.6; DB 13; Length 1503;

Best Local Similarity 51.3%; Pred. No. 2e-31;

Matches 657; Conservative 0; Mismatches 569; Indels 54; Gaps 5;

QY 301 TCGCGCGCTGTCATCCGCGAGCGCGCCATGCTACGAGCTGACACACCGGCTTCGGTCCCC 360

D 125 TCGCGGATATCGCGGAGCGCGGCGCGCTACGCGCATCAACACGGGCTTTGGGCGCC 184

QY 361 TTGCGAACCGCCCTGATCTCAGGTGAGATGTCGGAACGCTGCGAGCCCATCTTGTTCATC 420

Db 185 TCGCCAGCAGCATATTTCCGCA CGA CAA ACTGAACTGTTTGACGCGCAATCTGGTGTCT 244
QY 421 ATCTGGCCAGCGCGTGGGACCGGTGCTTGACTGGACGACGCGCGCGCATGGTCTCG 480
Db 245 CTACGGGTGCGGTGGGTGAGCCGATGTCGGTTCGGTCTGCTCTGATCGCG 304
QY 481 CGCGTCTGCTGATCGATCTCAGGGAGCCTCCGGTCCACGAGGGGACCATCGCTCGCC 540
Db 305 TGAAGACTCTCGAGCTCGCGCGCGCATTCGGGCGATTCGCCGTGAAGTGAATGACGCGC 364
QY 541 TGATCGACCTGCTCAATTCGAGCTCGCTCGGCGGTTCCAGCCGCGGACGAGTGGCG 600
Db 365 TGATCACTGCTGTACAAACGCGGACGCTGTCGGCTGATTTCCGCTCAAGGGTTCCGTCGGTG 424
QY 601 CGTCGGGTGACCTGACACGCGCTTTCGCGCATATGCTCTGCTCCACAGGGCGGGGAGACT 660
Db 425 CATCGGCGACCTTCGCGCCCTCGCGCATATGTCGGCGGCGCTGCTCGGCGTCGGCGA -- 482
QY 661 TCTTGGACCGGGACGGGACGCGCTTTGACGCGCGCAGAGGGCTCCGCGCGGACGGCTGC 720
Db 483 -AGTGTTCGGAAGGCGAGCGCATGCGGCGCACCGAGGGTCTCGCGCTCGTCGGCTCA 541
QY 721 AACCGCTCGATCTCTCCATCGCGATGCACTGGCGCTGTTCAACGGGACCTCCGCGCATGA 780
Db 542 AGCCGCTCACGCTGCAAGCCAAAGAGGGGCTGGCGCTCTCTGAAACGGCAGCGAGCTTCGA 601
QY 781 CGGGATCGCGCTGCTGAATGCTCAGCGCTGCGCGCATCTCGGCAACTGCGGCGTGGCGT 840
Db 602 CGCGCTCGCGCTGTACAAATGTTCCGCATCGAAGACCTGTACCGCACCGCGCTGGTGT 661
QY 841 TGACGCGCTCTGCTTGGGGAATGCTGAGAGGCGCGGACCGAGGCAATGGCGCGCGCACTGT 900
Db 662 CGGCGCATTTGTCGGTAGATGCCCAATGGCTCGGTCAAGCCGTTTCGACGCGCGCATTC 721
QY 901 CGAACTTGGCGCGCATTCGCGGACAGAGGACCGCGAGCGAGCTGCGGCGCGCGGTGG 960
Db 722 ACAGTTTGGCGCGCATCAAGGTCAGATCGACGCGCGCGCATACCGCTCGCTCTGCTG 781
QY 961 ACGGACGCGCGGTGCTCCGCGCTCATTTCCGCGGAGGCTCGACGCGCGGAYTA 1020
Db 782 AAGTTTCGGAATCAAGTTTTCGACG----- 808
QY 1021 TCGGGACGGACGGGCGGGGAGGATCCCTACAGCCTCGCTCGCTCGGCTCCGACGGTTC 1080
Db 809 -----CCGATTCCGACAGGTGACAGACCCGTACAGCTGCGTCCGACGCCAAGTCA 862
QY 1081 TCGGGCGGGCTTCGACAGCTCGCATGGCATACCGGCTGCTGACGATCGAGCTGAACG 1140
Db 863 TGGGCGGTGCTCTGGATCAGATCGGCCACGCGGCCAACGTGCTGCTCGAAGCGAACG 922
QY 1141 CGGTGACCGACATCCGCTGTTCCGCCGATGGGAGCGTGCCCGCTCGACGGGGGA 1200
Db 923 CGGTCTCCGACAACTCGCTGATTTTCCCGA-----CACCGGCGAAGTGTGTGCGGCGGTA 979
QY 1201 ATTTTCATGGGCCAGCATGTGGCGCTCAGCTCCGATCGCTCGCACGGCGTCAACCGTTC 1260
Db 980 ACTTCCACGAGACCGGTGCGCTTCGGCGCCGATTAACCTCGGCTCGCGCGCGCGAAA 1039
QY 1261 TGGCGGGCTTCGGGAGCGCCAGATTGCACTGTGACAGATGAAGGCTGAACCGTGGCG 1320
Db 1040 TCGGCGGTGGCGGAACGCGCATCGCTGCTGATCGACGCGACGCTGTGCGGCGTGC 1099
QY 1321 TGGCCCCCTTCCTCCACGGGGCCCCCGGGTTGAATTCGGCTTCATGGGGGACAGG 1380
Db 1100 CGCC-----GTTCTCTGTGCGGATGGCGGCGTGAATTCGGGCTTCATGATCGCGCAG 1153
QY 1381 TGACGGCGACCGGCT---CCTGGCGAGATGGAGCGACCGGACCTGCTCGATCCATT 1437
Db 1154 TCAGGCTGGGCGCTCGCTCGGAAAACNAGACGCTCGGCGATCCGGTTCGGTTCGATT 1213
QY 1438 CGATCTCCAGAACCGCGCAATCAGGATGTGCTTCGCTTGGGACCATCGCGCGCGCC 1497
Db 1214 CGTGGCCCACTTCGGCAAAACAGGAAGACCACTGTGATGGGACGTTTCGCGCGCGCA 1273

QY 1498 TCTGCGCGGAGAGATCGACCGTTGGCGGAGATCCTTCGATCCTCGTCTCTGCTTTG 1557
Db 1274 AGTTCGGCGACATTGCTGAAACACCGCAACATTCTGTGATCGAATGCTCGCGCGC 1333
QY 1558 CACAAGCTCGGAGCTGCGC 1577
Db 1334 CGCAAGGCTCGATCTGCGC 1353

RESULT 14
ACA25783

ID ACA25783 standard; DNA; 1521 BP.

XX ACA25783;

XX AC

XX 19-JUN-2003 (first entry)

XX Prokaryotic essential gene #7440.

XX Antisense; ds; prokaryotic essential gene; cell proliferation;

XX drug design; gene.

XX Burkholderia fungorum.

XX WO200277183-A2.

XX 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.

XX 06-SEP-2001; 2001US-00948993.

XX 25-OCT-2001; 2001US-0342923P.

XX 08-FEB-2002; 2002US-00072851.

XX 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX WPI; 2003-029926/02.

DR P-PSDB; ABU21913.

XX New antisense nucleic acids, useful for identifying proteins or screening

PT for homologous nucleic acids required for cellular proliferation to

PT isolate candidate molecules for rational drug discovery programs.

XX Claim 14; SEQ ID NO 13653; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of
the 6213 antisense sequences given in the specification where expression
of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC on a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for

identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate nucleic acids required for drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pat_sequences

Sequence 1521 BP; 263 A; 520 C; 480 G; 258 T; 0 U; 0 Other;

Query Match 11.4%; Score 211.6; DB 8; Length 1521;
Best Local Similarity 51.3%; Pred. No. 2e-31;
Matches 657; Conservative 0; Mismatches 569; Indels 54; Gaps 5;

301 TCGCGCTGTATCATCGGAGCGCGCCATGCTTACGGAATGCTGCAACCGGCTTCGGTCCCC 360
122 TCGCGATATCGCGCGAAGGGCGGCGCTACGGCATCAACACGGGCTTTGGGGCC 181
361 TTGGAAACCGGCTGATCTCAGGTGAGAAATGTCGAAACGCTGCGCAATCTTGTCCATC 420
182 TCGCAGCAGCATATTCGCGACGACCAACTCGAACTGTTGACGCGCAATCTGGTGCTCT 241
421 ATCTGGCCAGCGGCTGGACCGGTGCTTGAATGAGCAGCGCGCGCCATGTTCTGG 480
242 CTCACGGGTGGCGGTGGGAGCCGATGTCCGCTCGGTGCTGCTGATCGCGC 301
481 CGCGTCTGGTGTGATCGCTCAGGAGCGCTCCGCTGCGCAGCGGGGACCATCGCTCGCC 540
302 TGAACCTCTCGAGCTTCGGCGCGGCCATTCGGCAATTCGCCGTGAAGTATGACCGGC 361
541 TGATGACCTGCTCAATTCGAGCTCGCTCCGCGCTTCCAGCGCGCGGACCGTGGCG 600
362 TGATCAGCTGTATCAACCGCGACGTGCTGCGGTGATTCGGGTCAAGGGTTCCGTCGCTG 421
601 CGTCGGGTGACCTGACACCGCTTCCGATATGCTGCTGCTCCAGCGCGGGGAGACT 660
422 CATCGCGGACCTTCGCGCGCTCGCATATGTCGCGGGGCTGCTCGCGGTGCGCA-- 479
661 TCCTGACCGGAGCGGCTTACGCGCGGAGGCTCCGCGCGGAGCGCTG 720
480 -AGTGTTCGGAAGGCGGCGCATCGCGCCACCGAGGCTCTCGCTGTGCGCTCA 538
721 AACCGCTCGATCTTCCCATCGGATGCTGCGCTGCTCAACGGACCTCCGCCATGA 780
539 AGCGCTCAGCTGCAAGCAAGGCGCTGCGCTCCTGAAACGCGACGCGCTTCGA 598
781 CCGGGATCGCGTGTGATGCTCAGCGTCCGCCATCTCGGCAACTGGCGGTGGCGT 840
599 CGCGCTCGCGCTGTACAACATGTTCCGCAATCGAAGACCTGTACCGCACCGCGCTGCTGT 658
841 TGACGCGCTTGTGCGGAATGCTGAGAGCGCGGACCGAGGATGGCGCGCGCACTGT 900
659 CGGCGCATGTTGCGTATGATGCCAATGGGCTCGGTCAAGCCGTTTCGACGCGGCAATC 718
901 CCGACCTCGCGCGCTCCCGGACAGAGGACGCGCGGAGGCTCGCGCGCGCGCTGG 960
719 ACGAGTTGCGGCGCATCAAGTTCAGATCGACGCGCGCGGCGCATCGCTCGTGTG 778
961 ACGCAGCGCGCGGTGGTGGCGCACTGATTCGCGAGCGGAGGCTCGACCGCGCGGATA 1020
779 AAGGTTCCGCAATCAAGTTTCGCAAGC----- 805
1021 TCGGAGCGGAGCGGCGCGGCGGATGCTACAGCTGCGCTGCGCTCGCGAGTTC 1080
806 -----CCGATTGCGCAAGGTGCGAGACCGGTACGCTGCGCTGCGCGCAAGTCA 859
1081 TCGGCGCGGCTTCGACACGCTCGCATGAGTATGACCGGCTGCTGACGATGAGTGAACG 1140
860 TGGGCGGTGCTGTGATGATGATGCGCCACGCGGCGCAAGTGTGCTGCTCGAGCGAAG 919
1141 CGGTGACCGCAATTCGGTGTTCGCGCGCGATGCGAGCGTGGCGCGCTTCGACGGGGCA 1200

920 CGGTCTCGGACATCCGCTGATTTTCCCGA---CACGGGGAAGTGTCTCGGCGGTA 976
1201 ATTTTCATGGGCCAGCATGTGCGCTGAGTCCGATGCGCTCGCCACGCGCGCTCACCGTTC 1260
977 ACTTCCACGCGAGCGCGTTCGCTTCGCGGCGGATAACCTTCGCGCTCGCGCGCGCGAAA 1036
1261 TGGCGGCGCTTCGCGAGCGCGCATGTCAGTCTGACAGATGAAAGGCTGAACCGTGGGC 1320
1037 TCGCGCGCTTCGCGGAAACGCGCATCGCTGCTGATCGACGCGCATGTCGCGGCTTCG 1096
1321 TGCCCCCTTCTCCACGCGGCGCGCGCGGTTGAATTCGGCTTCATGGCGCGCACAGG 1380
1097 CGCC-----GTTCCTGTCGGCGATGCGCGGTGAATTCGGGCTTCATGATCGCGCAG 1150
1381 TGACGGCGACCGCGCT---CCTGCGCGAGATGCGAGCACGCGGACCTGCTCTGATTCATT 1437
1151 TCACGGCTCGCGGCTCGCTCGCGGAAACAGACGCTCGCGCATCCGCTTCGCTCGATT 1210
1438 CGATCTCCAGAACCGCGCAATCAGGATGTGCTGCTGCTGGGACCATCGCGCGGCGCC 1497
1211 CGTGGCCACTTCGGCAACACGAGAACACCGTGTGATGCGGACGTTTCGCGCGCGCA 1270
1498 TCTGCGCGAGAGATCGACCGTTGGCGGAGATCCTTGGGATCTCGCTCTGCTGTTG 1557
1271 AGTTCGGCGACATGCTGAAACACCGCGCAATTCCTGATCGAATGCTCGCGCGCG 1330
1558 CACAAGCTCGGAGCTCGCG 1577
1331 CGCAAGCGCTCGATCGCG 1350

RESULT 15
ADS50000
ID ADS50000 standard; cDNA; 1518 BP.
XX
AC ADS50000;
XX
DT 02-DEC-2004 (first entry)
XX
DE Bacterial polynucleotide #4743.
KW Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cell cycle pathway modification; plant growth regulator;
KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW bacterial polynucleotide; gene; ss.
OS Bacteria.
XX
XX US2003233675-A1.
XX
XX 18-DEC-2003.
XX
XX 20-FEB-2003; 2003US-00369493.
XX
XX 21-FEB-2002; 2002US-0360039P.
XX
XX (CAOY/) CAO Y.
XX (HINK/) HINKLE G J.
XX (SLAT/) SLATER S C.
XX (CHEN/) CHEN X.
XX (GOLD/) GOLDMAN B S.
XX
XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX WPI; 2004-061375/06.
XX
XX New recombinant DNA construct comprising a promoter positioned to provide
XX for expression of a polynucleotide encoding a polypeptide from a
XX microbial source, useful for producing plants with improved properties.

XX Claim 1; SEQ ID NO 28430; 122pp; English.
XX The invention relates to a recombinant DNA construct comprising a
CC promoter functional in a plant cell, where the promoter is positioned to
CC provide for expression of a polynucleotide encoding a polypeptide from a
CC microbial source. The invention also relates to a transformed plant
CC comprising the recombinant DNA construct and a method of producing a
CC transformed plant having an improved property. The plant is a crop plant
CC such as maize or soybean. The method of producing a transformed plant
CC having an improved property comprises transforming a plant with the
CC recombinant DNA construct and growing the transformed plant, where the
CC polynucleotide or polypeptide is useful for improving plant properties.
CC The recombinant DNA construct is useful for producing plants with
CC improved plant properties, e.g. improved cold, heat or drought tolerance,
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC increased resistance to plant disease, better growth rate by modification
CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or
CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition, improved lignin production or improved galactomannan
CC production. This sequence represents a bacterial polynucleotide used in
CC the scope of the invention. Note: The sequence data for this patent did
CC not form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 1518 BP; 254 A; 504 G; 504 G; 256 T; 0 U; 0 Other;

Query Match 9.6%; Score 177.4; DB 13; Length 1518;
Best Local Similarity 49.5%; Pred. No. 7.7e-25;
Matches 688; Conservative 0; Mismatches 631; Indels 72; Gaps 6;

QY 330 GTCTACGGACTGACACCGGCTTCGGTCCCTTCGGTAAACCGCTGATCTCAGGTGAGAA 389
DB 154 GTGTACGGGCTCAACACCGGTTACCGGATGCTTGGTGTGCTGACGCTGCGATGGAAC 213

QY 390 GTCCGAACGCTGAGCCCAATCTTGTCCATCATCTGCCAGCGGCTGGGACCGGTGCTT 449
DB 214 GTCCGAAGCACTGCGGTGACACTACCGGCTATACGGGTGCGGCAATGGGCCAGTATCTC 273

QY 450 GACTGGACGACGCGCGCGCATGGTCTTGGCGCGCTGCTGGTGTGCTCATCGCTCAGGGAGCC 509
DB 274 GACGAAGCGCAACGCTGCGGGTATCGCGCGCGGCTCAACTCGCTCGCATACGGTTT 333

QY 510 TCGGGTGCACGAGGGACCATCGCTCGCTGATCGACCTGCTCAATTCGAGCTCGCT 569
DB 334 TCGGGGCTGCTCGGTTGTCTCGAACGCTCTCGCGGATCTGGTCAATCATCGCGTGTG 393

QY 570 CCGGCGGTTCCAGCGCGACGCTGGGCGGTGCGGCTGCGGCTGACACCGCTTGGCAT 629
DB 394 CCGCGATTCGCTCGAAGGCTCGGTGCGGCGGAGCGGCGATCTCACGCCGCTTCGTAT 453

QY 630 ATGGTCTCTGCTCCAGGCGCGGGAGACTTCTCTGGACCGGGACGCGGCTTGAC 689
DB 454 GTGGCGCGCGCTCGCGGTGAGCGGAGCTTAATGTTCGAAGGGCAGTTGCGCAACGTG 513

QY 690 GGGCGAAGAGGGCTCGGCGCGGACCGGCTGCAACCGCTGATCTCTCCATCCGATGCA 749
DB 514 CGCGAGGTGTGAGACCGAACTCGGTCAAAACG--CCGCTCACGCTCGCGCGGAAAGGG 570

QY 750 CTGGCGCTGTTCAACGGGACCTCCGCCATGACCGGGATCGCGTGTGATGCTCAGGCC 809
DB 571 CTTCGCGCTGATGAACCGGACCGCGGTGATGACGGGTCTCGCCTGTCTCGGTTTGC GCG 630

QY 810 TGCGCGCATCTCGGCAACTGGGCGGTGGGCTTGACCGGCCCTGCTTCGGGAATGCTCTGAGA 869
DB 631 GCGGATCATCTACGGGCTCACCGCGCTGACCGGCTGTGCAAGCTCGCGCTCGAC 690

QY 870 GGGCGGACCGAGGATGGGCGCGGCACTGTGTCGAACCTGCGGCGCATCCCGGACAGAA 929
DB 691 GGGCGTGGCGCGCATTTTCGACGCAATGCTGTTTCGAGTGAAGCCGACCGCGCCAGGCC 750

QY 930 GACGCCGACGAGGCTGCGCGCCGCTGTGGACGGCAGCGCGCGGTGTTCGGCAGTC 989
DB 751 GAAGCGCGCATGTGATCCGCGACGACCTTGGCGGCGCGA----- 791

QY 990 ATTGCCGACGGAGGCTGACCGCGCGGATATCGGACGAGCGGAGGGGGGAGGAT 1049
DB 792 -----CGACACGCGCGGCGCACCGCTTTTCAGGAC 819

QY 1050 GCTACAGCTGCGCTGCGCTCCGAGGTTCTCGGGGCGGCTTCGACACGCTCGCATGG 1109
DB 820 CGTATTTCGATTGCTGCGCGCGCGACGATCGGCTGCGCGCGACCGGCTTCGTGG 879

QY 1110 CATGACCGGCTGTGCTGAGATCGAGCTGAACGGGCTGACCGCAATCCGGTGTTCGCGCCC 1169
DB 880 GTACGCGCGGACATCGAGAAACGAACTGAACAGCGCGCAACCGCTGATCGATCCG 939

QY 1170 GATGGAGGTCGCCGCCCTTCGACCGGGGCAATTTTCATGGGCGAGATGTGGCGCTGACG 1229
DB 940 GA---CAACGAAACGGCTGTCTGACCGCGCGCACTTCTACGGCGGTTCATATCGCTTCGCG 996

QY 1230 TCCGATGCGCTCGCCACGCGCTGACCGTTCTGGGCGGCTTTCGAGCGCCAGATTGCA 1289
DB 997 ATGGACTCGCTGAAGTTCGCGGTGCGCAACCTCGCGGATCTGATGACCGGCACTGGCG 1056

QY 1290 CGTCTGACAGATGAAGGCTGAACCGTGGGCTGCCCTTCTCCACCG-----G 1340
DB 1057 CTCTCTGCTGACGTCAATTTCAATAACGGCTTGGCGCGCAATCTGTTCGGGCGCTACGTCC 1116

QY 1341 GGGCCCGCGGGTTGAATTCGCGCTTCATGGGCGCACAGGTGACGGCGACCGCGCTCCTG 1400
DB 1117 GCGCGCGCGCATCAATCAACGGGTTCAAGGCGGTGACAGATTTCTCTCATCCGATGAGACC 1176

QY 1401 GCGGA---GATGCGAGCCACGGGACCTGCTCGATTCATTCGATCTCCAGGACGCGGCC 1457
DB 1177 GCGGAAGCATTTGAAGAACCAACATGCGCGCGAGCGTTTCTCGCGCTCCACCGAGGCGCAC 1236

QY 1458 AATCAGATGTGCTCTGCTTGGGACCATCGCGCGCGCTCTGCGCGCAGAAAGATCGAC 1517
DB 1237 ACCAGGACAAAGTCAGTCAGTCAGGACGATCGCGCGCGGAGTCTCTGCGCTGCTGGA 1296

QY 1518 CGTTGGGCGGAGATCCTTCGCGATCTCTGCTCTCTGTTGCAACAAGCTCGGAGCTGCGC 1577
DB 1297 CTGACCGAAACAGTGGCGGCGAGCGCATACGCTCGCAACCGTGCAGCTGCCGCTCTGCGA 1356

QY 1578 TCGGGCAGCG---GCCTAGACGGGGTGTCTCCGCGGGGGAAGAGCTGTGCGAGGCGCTG 1634
DB 1357 TTGAAGATCGATAGCGGAAACACCGGTTCCCGCGCGCTCGGAACGTTTCATGAGAGCGTG 1416

QY 1635 CGCGAGCAGTTCCCGCGCTTGAGACGAGCCGCGCCCTCGGACAGGAAATTCGCGCGCTT 1694
DB 1417 AGCGCGCAGTTCGCGGTTCTGTGACGAGACCGCGGCTCGAACGACATGCGCGCCCTG 1476

QY 1695 GCTACGCACT 1705
DB 1477 ACCGCGCGCAT 1487

Search completed: December 10, 2005, 12:35:04
Job time : 1140 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2005, 09:06:26 ; Search time 186 Seconds
(without alignments)
1242.544 Million cell updates/sec

Title: US-10-621-826-3

Perfect score: 2668

Sequence: 1 VKPMLANSPRPKPAVELDRHI.....RPLQGEIATLALHLLQQSPV 526

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A: Geneseq 21:*

- 1: geneseqp1980s:*
- 2: geneseqp1990s:*
- 3: geneseqp2000s:*
- 4: geneseqp2001s:*
- 5: geneseqp2002s:*
- 6: geneseqp2003as:*
- 7: geneseqp2003bs:*
- 8: geneseqp2004s:*
- 9: geneseqp2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2668	100.0	526	8 ADJ57136	Adj57136 R. sphaer
2	2444	91.6	482	8 ADN25163	Adn25163 Bacterial
3	1335	50.0	539	8 ADL16888	Adl16888 Rhodobact
4	1335	50.0	540	8 ADL16877	Adl16877 Rhodobact
5	1333	50.0	540	8 ADL16887	Adl16887 Rhodobact
6	837	31.4	529	8 ADN26159	Adn26159 Bacterial
7	814.5	30.5	539	3 AAB13586	Aab13586 Streptomy
8	761	28.5	508	9 ABM95993	Abm95993 M. xanthu
9	755	28.3	501	8 ADN22103	Adn22103 Bacterial
10	755	28.3	501	8 ADN24862	Adn24862 Bacterial
11	755	28.3	507	6 ABU21913	Abu21913 Protein e
12	753.5	28.2	501	8 ADS30178	AdS30178 Bacterial
13	749.5	28.1	507	6 ABU22102	Abu22102 Protein e
14	748	28.0	522	8 ADN26467	Adn26467 Bacterial
15	734.5	27.5	502	8 ADN24391	Adn24391 Bacterial
16	734.5	27.5	511	6 ABU21162	Abu21162 Protein e
17	729.5	27.3	487	8 ADN21635	Adn21635 Bacterial
18	729	27.3	460	8 ADS21195	AdS21195 Bacterial
19	728.5	27.3	508	8 ADS44935	AdS44935 Bacterial
20	715	26.8	494	8 ADS24598	AdS24598 Bacterial
21	712	26.7	511	8 ADS28241	AdS28241 Bacterial
22	711.5	26.7	524	8 ADS45022	AdS45022 Bacterial
23	709	26.6	506	8 ADN24850	Adn24850 Bacterial
24	709	26.6	506	8 ADN22090	Adn22090 Bacterial

25	708.5	26.6	511	6 ABU49321	Abu49321 Protein e
26	706	26.5	496	8 ADS25085	AdS25085 Bacterial
27	703.5	26.4	485	8 ADS21886	AdS21886 Bacterial
28	699.5	26.2	484	8 ADN25988	Adn25988 Bacterial
29	695.5	26.1	657	8 ADP24794	Adp24794 PRO polyp
30	695	26.0	510	8 ADS22467	AdS22467 Bacterial
31	694.5	26.0	657	8 ADN04255	Adn04255 Antipsoi
32	694.5	26.0	657	8 ADP23988	Adp23988 PRO polyp
33	694	26.0	497	8 ADN24930	Adn24930 Bacterial
34	688.5	25.8	511	6 ABU16504	Abu16504 Protein e
35	688	25.8	507	8 ADS26381	AdS26381 Bacterial
36	688	25.8	507	8 ADS27132	AdS27132 Bacterial
37	688	25.8	507	8 ADS26749	AdS26749 Bacterial
38	687	25.7	482	8 ADN22171	Adn22171 Bacterial
39	687	25.7	525	8 ADL05169	AdL05169 M. catar
40	683.5	25.6	491	4 AAU34175	Aau34175 Staphyloc
41	683.5	25.6	504	4 AAU37099	Aau37099 Staphyloc
42	683.5	25.6	556	4 AAU43358	Aau43358 Propionib
43	683.5	25.6	556	6 ABM39877	Abm39877 Propionib
44	683	25.6	508	7 ABO67058	AbO67058 Klebsiell
45	683	25.6	519	6 ABU35328	Abu35328 Protein e

ALIGNMENTS

RESULT 1
ADJ57136
ID ADJ57136 standard; protein; 526 AA.
XX
AC ADJ57136;
XX
DT 06-MAY-2004 (first entry)
XX
DE R. sphaeroides tyrosine ammonia lyase (TAL) enzyme.
XX
KW TAL; tyrosine ammonia lyase; catalyst; para-hydroxycinnamic acid; PHCA;
KW tyrosine; enzyme.
XX
OS Rhodobacter sphaeroides.
XX
PN WO2004009795-A2.
XX
PD 29-JAN-2004.
XX
PF 23-JUL-2003; 2003WO-US023229.
XX
PR 23-JUL-2002; 2002US-0397820P.
XX
FA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Huang L, Xue Z;
XX
DR WPI; 2004-214331/20.
XX
DR N-PSDB; ADJ57135.
XX
PT Novel tyrosine ammonia lyase enzyme, used as catalyst in producing para-hydroxycinnamic acid from tyrosine.
XX
PS Claim 5; SEQ ID NO 3; 53pp; English.
XX
CC The invention relates to a tyrosine ammonia lyase enzyme (TAL) and encoding polynucleotides. The TAL polypeptide is used as a catalyst in producing para-hydroxycinnamic acid (PHCA) from tyrosine. The method involves contacting a transformed host cell expressing the polypeptide with a fermentable carbon substrate, e.g. monosaccharides, oligosaccharides, polysaccharides, carbon dioxide, methanol, formaldehyde, formate and carbon-containing amines, preferably glucose, growing the host cell for a time sufficient to produce PHCA, and optionally recovering PHCA. The TAL polypeptide enables simple and cost-effective production of PHCA. The present sequence represents a R. sphaeroides TAL polypeptide.
XX

QY 463 LAILALCLAAAEELRCGSLDGVSPAGKXIVQALREQFPFLETRDPLGQEIATLATHLQ 522
DB 421 LAILALCLAAAEELRCGSLDGVSPAGKXIVQALREQFPFLETRDPLGQEIATLATHLQ 480
QY 523 QS 524
DB 481 QS 482

RESULT 3
ADL16888
ID ADL16888 standard; protein; 539 AA.
AC ADL16888;
XX
DT 06-MAY-2004 (first entry)
DE Rhodobacter capsulatus TAL mutant protein.
KW Para-hydroxycinnamic acid; photoactive yellow protein;
KW photochromic substance; biochip; liquid crystal polymer; PYP; enzyme;
KW mutant.
XX
OS Rhodobacter capsulatus.
XX Synthetic.
XX
PN US2004029230-A1.
XX
PD 12-FEB-2004.
XX
PF 18-JUN-2003; 2003US-00464609.
XX
PR 18-JUN-2002; 2002US-0389593P.
XX
PA (KYND/) KYNDT J J A.
PA (BEEU/) BEEUMEN J V.
XX
PI Kyndt JJA, Beeumen JV;
XX WPI; 2004-168893/16.
XX
XX New nucleic acid encoding tyrosine-ammonium lyase polypeptide, for
PT synthesizing para-hydroxycinnamic acid and holo-photoactive yellow
PT protein used as photochromic substance in electro optical random access
PT memory or in biochips.
XX
PS Example 1; Page; 32pp; English.
XX
CC The present invention is generally related to recombinant DNA technology
CC and particularly to DNA strands useful for the production of
CC parahydroxycinnamic acid and photoactive yellow protein (PYP) in a
CC suitable host expression system. The invention is useful for synthesizing
CC parahydroxycinnamic acid and derivative products such as holo-photoactive
CC yellow protein which is useful as a photochromic substance in electro
CC optical random access memory, in biochips, in light-inducible gene
CC expression systems and in liquid crystal polymers. The present sequence
CC is Rhodobacter capsulatus tyrosine ammonia lyase (TAL) mutant protein.
CC The mutant protein is constructed with the deletion of Ala at position
CC 535. Note: This protein is shown in the specification but is derived from
CC Rhodobacter capsulatus wild type tyrosine ammonia lyase (TAL) protein
CC shown as SEQ ID NO 4 in page 13-14 of this specification.
XX
SQ Sequence 539 AA;

Query Match 50.0%; Score 1335; DB 8; Length 539;
Best Local Similarity 56.5%; Pred. No. 2.1e-105;
Matches 292; Conservative 59; Mismatches 162; Indels 4; Gaps 4;

QY 11 KPAVELDRHIDQAHAVASGARTVLAPPARDRCRASEARLIGAVIREARHYGLTTGFG 70
DB 19 KDCIALDGLATLVQCEATATHRISRTVPALRRCARAHAEARHIAEQRIHYGTTGFG 78
QY 71 PLANRLISGENVRTLQANLVHHLASGVGVLDWTTRAMVRLARLVSIAGGASGASEGTIA 130

DB 79 PLANKLIGADQAEIQQLNLIYHLATGVGPXKLSWAERALMLARLNSIIQGASGASPETID 138
QY 131 RLIDILNSELAPAVPSRGTVGASGDLTFLAHMVLCLQGRGDFLDRDGTGLDCAEGLRRGR 190
DB 139 RIVAVLNAGFAPEVPAQGTVGASGDLTFLAHMVLALQGRGMIDPSGRVQVEAGAVMDRLC 198
QY 191 LQPLDLSHRDALALVNGTSAMTGI-ALVNAHACRHNGWVALTALLAECLRGRTEANAA 249
DB 199 GGPLTLAARDGLALVNGTSAMTATAALTGVEAARAI-DAALRHSAVLMEVLSGHAEAWHP 257
QY 250 ALSDLRPHPGOKDRAARLRARVDGSRVVRVIAERRLDAGDIGTEPEAGODAYSIRCAP 309
DB 258 AFAEURPHPGQURATERLAQALDAGRVCRILTAAARLTAADLRDEHPAQDAYSLRVVP 317
QY 310 QVLGAGFDTLAWHDIRVLTIELNAVTDNFPFPGDSVPALHGNFNGHQHVALTSDALATAV 369
DB 318 QLVGAVMTLDWHDRVVTCELNSVTDNPIFPEGCAVPALHGNFNGHVHVALASDALNAAL 377
QY 370 TVLAGLABRQIARLTDERLNKGLPPFLHRGPAGLNSGFMGAQVTTATALLAENRATG-PAS 428
DB 378 VTLAGLVERQIARLTDEKLNKGLPAFLHGGQAGLQSGFMGAQVTTATALLAENRANATPVS 437
QY 429 IHSISTNAANODVVSGLGTIAARLCREKIDRWAEILAILALCLAAAEELRCG-SGLDGVSP 487
DB 438 VQSLSTNGANQDVVSMGTIAARRAQLPLSQIAILALALAAQAWDLDDPEQGQWSL 497
QY 488 AGKKLVQALREQFPFLETRDPLGQEIATLATHLQQS 524
DB 498 TARDLRDRIRAVSPGLRADRPLAGDIEAQAQGLRHPS 534

RESULT 4
ADL16877
ID ADL16877 standard; protein; 540 AA.
XX
AC ADL16877;
XX
DT 06-MAY-2004 (first entry)
XX
DE Rhodobacter capsulatus tyrosine ammonia lyase (TAL) protein.
KW Para-hydroxycinnamic acid; photoactive yellow protein;
KW photochromic substance; biochip; liquid crystal polymer; PYP; enzyme.
XX
OS Rhodobacter capsulatus.
XX
FH Key Location/Qualifiers
FT Misc-difference 534..535 /note= "Encoded by TCCNNNGCC"
XX
US2004029230-A1.
XX 12-FEB-2004.
XX 18-JUN-2003; 2003US-00464609.
XX 18-JUN-2002; 2002US-0389593P.
PA (KYND/) KYNDT J J A.
PA (BEEU/) BEEUMEN J V.
XX
XX Kyndt JJA, Beeumen JV;
XX
DR WPI; 2004-168893/16.
DR N-PSDB; ADL16876.
XX
PT New nucleic acid encoding tyrosine-ammonium lyase polypeptide, for
PT synthesizing para-hydroxycinnamic acid and holo-photoactive yellow
PT protein used as photochromic substance in electro optical random access
PT memory or in biochips.
XX
PS Claim 23; SEQ ID NO 4; 32pp; English.

XX The present invention is generally related to recombinant DNA technology
CC and particularly to DNA strands useful for the production of
CC parahydroxycinnamic acid and photoactive yellow protein (pyp) in a
CC suitable host expression system. The invention is useful for synthesizing
CC parahydroxycinnamic acid and derivative products such as holo-photoactive
CC yellow protein which is useful as a photochromic substance in electro
CC optical random access memory, in biochips, in light-inducible gene
CC expression systems and in liquid crystal polymers. The present sequence
CC is Rhodobacter capsulatus tyrosine ammonia lyase (TAL) protein.
XX
SQ Sequence 540 AA;

Query Match 50.0%; Score 1335; DB 8; Length 540;
Best Local Similarity 56.5%; Pred. No. 2.1e-105;
Matches 292; Conservative 59; Mismatches 162; Indels 4; Gaps 4;

Qy 11 KPVELDRHIDLDQAHAVASGGARIVLAPPARDRCRSEARLGAIVREARHVGLTTGFG 70
Db 19 KDCIALDGLTLVQCEAIATHRSRISVTPALRERCARAHARLEHAIAEQRHIYGIITGFG 78
Qy 71 PLANRLISGENVRTLOANLVHILASGVGVLDWTTARAVLRLVSLAQGASGASGECTIA 130
Db 79 PLANRLIGADQGAELQONLIYHLATGVGPKLSWAEARALMLRLNSILOGASGPETID 138
Qy 131 RLIDLNSLSELPAPVPSRGTVGASGDLTPLAHMVLCLQGRGDFLDRDGTDLGAEGLRRGR 190
Db 139 RIVAVLNAGFAPEVPAQGTGASGDLTPLAHMVLALQGRGMIDPSGRVQEGAVMDRLC 198
Qy 191 LQPLDLSHRDALALVNGTSAMTGI-ALVNAHACRHLGNVAVALTALLAECLRGRTAWAA 249
Db 199 GGPLETLAARDGLALVNGTSAMTAAITGVEAARAI-DAALRHSAVLMVEVLSGHAERAWHP 257
Qy 250 ALSDLRPHPGQKDAARLARVDGSAVRVHVTAERELDAGDGTGTEPAGQDAYSRLCAP 309
Db 258 AFALRPHPGQLRATERLAQALDAGAGVCRVTLTAARLTAADLRPEDHPAQDAYSRLVVP 317
Qy 310 QVLGAGFDTLAWHDRVLTIELNAVTNPVPPDGSPALHGNFMGOHVALTSDALATAV 369
Db 318 QLVGAVWDTLDWHDVVTCELNSVTNPPIPEGCAVPAHLHGNFMGVHVALASDALNAAL 377
Qy 370 TVIAGLAERQIARLTDERLNRGLPPFLHRGPAGLNSGFMGAQVTTATALLAEMRATG-PAS 428
Db 378 VTLAGLVERQIARLTDEKLNKGLPAFLHGGQAGLQSGFMGAQVTTATALLAEMRANATPVS 437
Qy 429 IHSISTNAANQDVVSLGTTIARLCKREKIDRWABILAILALCLAAQAEELRCG-SGLDGVSP 487
Db 438 VQSLSTNGANQDVVSMGTIARRARAQQLPLSQIQAILALALAAQAMDLLDDPPEGQWLSL 497
Qy 488 AGKKLVQALREQFPFLETDRPLGQETAAALATHLIQQS 524
Db 498 TARDLRRIIRAVSGLRADRPPLAGDIEAVAGLRRHPS 534

RESULT 5

ADL16887
ID ADL16887 standard; protein; 540 AA.
XX
AC ADL16887;
XX
DT 06-MAY-2004 (first entry)
XX
DE Rhodobacter capsulatus TAL mutant protein (D522H).
XX
KW Para-hydroxycinnamic acid; photoactive yellow protein;
KW photochromic substance; biochip; liquid crystal polymer; pyp; enzyme;
KW muten; mutant.
XX
OS Rhodobacter capsulatus.
OS Synthetic.
XX
PH Key Location/Qualifiers
FT Misc-difference 522

FT
XX
PN
XX
PD
XX
PF
XX
PR
XX
PA
PA
XX
XX
PI
XX
DR
XX
PT
PT
PT
XX
PS
XX
CC

US2004029230-A1.

12-FEB-2004.

18-JUN-2003; 2003US-00464609.

18-JUN-2002; 2002US-0389593P.

(KYND/) KYNDT J J A.

(BEEU/) BEEUMEN J V.

Kyndt JJA, Beeumen JV;

WPI; 2004-168893/16.

New nucleic acid encoding tyrosine-ammonium lyase polypeptide, for
synthesizing para-hydroxycinnamic acid and holo-photoactive yellow
protein used as photochromic substance in electro optical random access
memory or in biochips.

Example 1; Page; 32pp; English.

The present invention is generally related to recombinant DNA technology
and particularly to DNA strands useful for the production of
parahydroxycinnamic acid and photoactive yellow protein (pyp) in a
suitable host expression system. The invention is useful for synthesizing
parahydroxycinnamic acid and derivative products such as holo-photoactive
yellow protein which is useful as a photochromic substance in electro
optical random access memory, in biochips, in light-inducible gene
expression systems and in liquid crystal polymers. The present sequence
is Rhodobacter capsulatus tyrosine ammonia lyase (TAL) mutant protein.
Note: This protein is shown in the specification but is derived from
Rhodobacter capsulatus wild type tyrosine ammonia lyase (TAL) protein
shown as SEQ ID NO 4 in page 13-14 of this specification.

SQ Sequence 540 AA;

Query Match 50.0%; Score 1333; DB 8; Length 540;

Best Local Similarity 56.5%; Pred. No. 3.2e-105;

Matches 292; Conservative 58; Mismatches 163; Indels 4; Gaps 4;

Qy 11 KPVELDRHIDLDQAHAVASGGARIVLAPPARDRCRSEARLGAIVREARHVGLTTGFG 70
Db 19 KDCIALDGLTLVQCEAIATHRSRISVTPALRERCARAHARLEHAIAEQRHIYGIITGFG 78
Qy 71 PLANRLISGENVRTLOANLVHILASGVGVLDWTTARAVLRLVSLAQGASGASGECTIA 130
Db 79 PLANRLIGADQGAELQONLIYHLATGVGPKLSWAEARALMLRLNSILOGASGPETID 138
Qy 131 RLIDLNSLSELPAPVPSRGTVGASGDLTPLAHMVLCLQGRGDFLDRDGTDLGAEGLRRGR 190
Db 139 RIVAVLNAGFAPEVPAQGTGASGDLTPLAHMVLALQGRGMIDPSGRVQEGAVMDRLC 198
Qy 191 LQPLDLSHRDALALVNGTSAMTGI-ALVNAHACRHLGNVAVALTALLAECLRGRTAWAA 249
Db 199 GGPLETLAARDGLALVNGTSAMTAAITGVEAARAI-DAALRHSAVLMVEVLSGHAERAWHP 257
Qy 250 ALSDLRPHPGQKDAARLARVDGSAVRVHVTAERELDAGDGTGTEPAGQDAYSRLCAP 309
Db 258 AFALRPHPGQLRATERLAQALDAGAGVCRVTLTAARLTAADLRPEDHPAQDAYSRLVVP 317
Qy 310 QVLGAGFDTLAWHDRVLTIELNAVTNPVPPDGSPALHGNFMGOHVALTSDALATAV 369
Db 318 QLVGAVWDTLDWHDVVTCELNSVTNPPIPEGCAVPAHLHGNFMGVHVALASDALNAAL 377
Qy 370 TVIAGLAERQIARLTDERLNRGLPPFLHRGPAGLNSGFMGAQVTTATALLAEMRATG-PAS 428
Db 378 VTLAGLVERQIARLTDEKLNKGLPAFLHGGQAGLQSGFMGAQVTTATALLAEMRANATPVS 437
Qy 429 IHSISTNAANQDVVSLGTTIARLCKREKIDRWABILAILALCLAAQAEELRCG-SGLDGVSP 487
Db 438 VQSLSTNGANQDVVSMGTIARRARAQQLPLSQIQAILALALAAQAMDLLDDPPEGQWLSL 497

PA (SLAT/) SLATER S C.
PA (CHEN/) CHEN X.
PA (GOLD/) GOLDMAN B S.
XX
XX
XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX WPI; 2004-061375/06.
XX
XX New recombinant DNA construct comprising a promoter positioned to provide
PT for expression of a polynucleotide encoding a polypeptide from a
PT microbial source, useful for producing plants with improved properties.
XX
XX Claim 1; SEQ ID NO 7515; 122pp; English.
XX
XX The invention relates to a recombinant DNA construct comprising a
CC promoter functional in a plant cell, where the promoter is positioned to
CC provide for expression of a polynucleotide encoding a polypeptide from a
CC microbial source. The invention also relates to a transformed plant
CC comprising the recombinant DNA construct and a method of producing a
CC transformed plant having an improved property. The plant is a crop plant
CC such as maize or soybean. The method of producing a transformed plant
CC having an improved property comprises transforming a plant with the
CC recombinant DNA construct and growing the transformed plant, where the
CC polynucleotide or polypeptide is useful for improving plant properties.
CC The recombinant DNA construct is useful for producing plants with
CC improved plant properties, e.g. improved cold, heat or drought tolerance,
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC increased resistance to plant disease, better growth rate by modification
CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or
CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition, improved lignin production or improved galactomannan
CC production. This sequence represents a bacterial polypeptide used in the
CC scope of the invention. Note: The sequence data for this patent did not
CC form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.
XX
XX SQ Sequence 501 AA;

Query Match 28.3%; Score 755; DB 8; Length 501;
Best Local Similarity 38.8%; Pred. No. 1.1e-55;
Matches 199; Conservative 73; Mismatches 211; Indels 30; Gaps 11;
QY 19 HIDLQAHAVASGARIVLAP---PARDRCSEARLGVIREARHVYGLTTCGGLANR 75
DB 9 HLTLPQQRQAREHVALQDLPASHAAIDACAQAVADIAA---KGEPAYGINTGFRLAST 65
QY 76 LISGVNRTLQANLVHHLASGVGVPLDWTTRAMVLARLVSIQAQASGEGTARLIDL 135
DB 66 HIPHQLELLQRLNLVSHAVGVGEPMSRPVVRLLIALKLSLGRHSGIRREVMDALITL 125
QY 136 LNSELAPVPSRGTVGASGDLTPLAHMVLCLQGRGDFLDRDGTLDGAEGLRRGRLOPLD 195
DB 126 YNADVLVPIVKGSVGSAGDLPLAHMSAALLGVGEVFAK-GERMPATEGLVALGLKPLT 184
QY 196 LSHRDALAVNGTSMTGIALVNAHACRHLGNVAVALTALLAECLRGRTAWAALSLDR 255
DB 185 LQAKGGLALNGTQASTALAYNMFAIEDLYRTALVSGALSVDAAMGSKVPFDPARHLEL 244
QY 256 PHPGQKDAARLARVDGSRVVRVHRAERRLDAGDGTGTEPEAGQAYSRLCAPOVLGAG 315
DB 245 GHQQIDAAAYRSLLEGSAINVSHADCK------VDPPYSRLCQPOWQAC 291
QY 316 FDTLAMDRLVLTTELNAVTPNP-VFPDPGVPALHGGNFQGVHVALTSDALATAVTVLAG 374
DB 292 LDQWRHAANVLLLEFANAVSDNPLIFPDGTGV--LSGNGFHAEPVFAADNLALAAAEIGA 349
QY 375 LAERQIARLDERNLNGLPPTFLHRGPAAGLNSGFGAQVTTATALLAEMRATG-PASIHIS 433
DB 350 LAERRIALLIDATLUS-GLPPFLVR-DGVNGSVGFMAHVTTAALASENKTLAHPASVDSLP 407

QY 434 TNAANQDVVSLGTIAARLCREKIDRWAEILAILALCLQAARLRCGSLDGVSPAGKKLV 493
DB 408 TSANQEDHVSMTAFFAARKLGDIAENTANILLSELLAAQGVDLRAP---HKTSPSLQKAM 464
QY 494 QALREQFPFPLETORPLGQBIATAALATHLLQOSPV 526
DB 465 DTVRKDVHAYELDPFAPDIAAV-TRLVQNGTI 496
RESULT 11
ABU21913
ID ABU21913 standard; protein; 507 AA.
XX
XX AC ABU21913;
XX
XX DT 19-JUN-2003 (first entry)
XX
XX DE Protein encoded by Prokaryotic essential gene #7440.
XX
XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
XX OS Burkholderia fungorum.
XX
XX FN W0200277183-A2.
XX
XX PD 03-OCT-2002.
XX
XX PF 21-MAR-2002; 2002WO-US009107.
XX
XX PR 21-MAR-2001; 2001US-00815242.
XX
XX PR 06-SEP-2001; 2001US-00948993.
XX
XX PR 25-OCT-2001; 2001US-0342923P.
XX
XX PR 08-FEB-2002; 2002US-00072851.
XX
XX PR 06-MAR-2002; 2002US-0362699P.
XX
XX PA (ELIT-) ELITRA PHARM INC.
XX
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
XX WPI; 2003-029926/02.
XX
XX N-PSDB; ACA25783.
XX
XX PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 25; SEQ ID NO 49837; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational

[illegible]

KW homologous recombination; seed oil yield; protein yield; carbohydrate; lignin; galactomannan; bacterial polypeptide.

OS Bacteria.

PN US2003233675-A1.

PD 18-DEC-2003.

XX 20-FEB-2003; 2003US-00369493.

PF 21-FEB-2002; 2002US-0360039P.

PR (CAO Y.) CAO Y.

XX (HINKLE G J.) HINKLE G J.

PA (SLATER S C.) SLATER S C.

PA (CHEN X.) CHEN X.

PA (GOLDMAN B S.) GOLDMAN B S.

XX

PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;

XX WPI; 2004-061375/06.

DR

XX

XX New recombinant DNA construct comprising a promoter positioned to provide for expression of a polynucleotide encoding a polypeptide from a microbial source, useful for producing plants with improved properties.

PT

PS Claim 1; SEQ ID NO 9120; 122pp; English.

XX

XX The invention relates to a recombinant DNA construct comprising a promoter functional in a plant cell, where the promoter is positioned to provide for expression of a polynucleotide encoding a polypeptide from a microbial source. The invention also relates to a transformed plant comprising the recombinant DNA construct and a method of producing a transformed plant having an improved property. The plant is a crop plant such as maize or soybean. The method of producing a transformed plant having an improved property comprises transforming a plant with the recombinant DNA construct and growing the transformed plant, where the polynucleotide or polypeptide is useful for improving plant properties. The recombinant DNA construct is useful for producing plants with improved plant properties, e.g. improved cold, heat or drought tolerance, increased resistance to herbicides, extreme osmotic conditions, pathogens or pests, increased resistance to plant disease, better growth rate by modification of the cell cycle pathway with plant growth regulators, increased rate of homologous recombination, modified seed oil or protein yield and/or content, improved yield by modification of carbohydrate, nitrogen or phosphorus use and/or uptake, by modification of photosynthesis or by providing improved plant growth and development under at least one stress condition, improved lignin production or improved galactomannan production. This sequence represents a bacterial polypeptide used in the scope of the invention. Note: The sequence data for this patent did not form part of the printed specification but was obtained in electronic format from USPTO at seqdata.uspto.gov/sequence.html.

XX

SQ Sequence 522 AA;

Query Match

Best Local Similarity 28.0%; Score 748; DB 8; Length 522;

Matches 205; Conservative 70; Mismatches 205; Indels 46; Gaps 13;

QY 20 IDLQAHAVASG--GARTV-LAPPARDRCRASEARLGVAREARHVVGLTTGFGFLANRL 76

DB 8 LTIEQVLAVAGQPTVPVRLAPARQVRERAQVQDLLARGVVAYGITTGFGAFKDRV 67

QY 77 ISGENVRLQANLVHVLASGVPVLDWTARAVLRLVSTAGASGASEGTIARLDLL 136

DB 68 IAPQVERLQNLVSHVAVGVPVFDIPTTIRAILIRANTLARGHSGVRLQVTERLDML 127

QY 137 NSELAPVPSRTGTCASGDPLAHMVLCLQGRGDFLDRDCTLDGAGLGRGLQPLDL 196

DB 128 NQGIHPRIPCKGSLGASGDPLAHMALPLIGLGE-VWQGEVLPAATALERLQWQPUHL 186

QY 197 SHRDALALVNGTSAMTGIALVNAHACRHLGNWALVATALLAECLRGRTTEAALSDLRP 256

DB 187 AKEGLALTNGTAVWCALGVITETARETSLSATDIAGCLSLAEALYGTTPAEDARLHALRP 246

QY 257 HPGQKDAARLARVRDGSARVVRHVIAERRLDAGDIGTEPEAGQDAYSLRCAPQVLGAGF 316

DB 247 FPRQIECAHLRLLAGST-FVRN-----NDPRHVQDAYTLRCIPQVHGAVR 292

QY 317 DTLAWHRVLTIELNAVTDNVPF--PPDGSVPALHGGNFMGQHVALTSDALATATVVLG 374

DB 293 DAIAAYARWVFAIELNAVTDNPLLFVDDGNGVEVSGGNFHGPELAIALDYLGLAVAEIGN 352

QY 375 LAERQIARLTDERLN-RGLPPFLHRGPAGLNSGPMGAQVATALLAEARATG-PASIHSI 432

DB 353 IAEERLWRLTDEASNTVLPALFTRA--GGLNSGMFVVOYTAAALATENKVLAHPASVDSI 411

QY 433 STNAANQDVVSLGTIAARLCREKIDRWAEIILALCLAQAAELR---CGSGL----DGVS 486

DB 412 PTSANVEDHVMGVTAGLKLRSIIDNVSQIILALELFAAAQGIFFRQELSGQARLGRGTG 471

QY 487 PAGKKLVQALREQPPPLETRPL-----GQETIALLATH 519

DB 472 P---VVLEIRQYVFFIAEDTLLHPYITIISELVAQGGKIAAAAAVH 513

RESULT 15

ADN24391

ID ADN24391 standard; protein; 502 AA.

XX

AC ADN24391;

XX 02-DEC-2004 (first entry)

DE Bacterial polypeptide #7044.

XX

KW Recombinant DNA construct; transformed plant; improved plant property; cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis; pathogen tolerance; pest tolerance; plant disease resistance; cell cycle pathway modification; plant growth regulator; homologous recombination; seed oil yield; protein yield; carbohydrate; nitrogen; phosphorus; photosynthesis; lignin; galactomannan; bacterial polypeptide.

XX

OS Bacteria.

XX US2003233675-A1.

XX 18-DEC-2003.

XX 20-FEB-2003; 2003US-00369493.

XX 21-FEB-2002; 2002US-0360039P.

PA (CAO Y.) CAO Y.

PA (HINKLE G J.) HINKLE G J.

PA (SLATER S C.) SLATER S C.

PA (CHEN X.) CHEN X.

PA (GOLDMAN B S.) GOLDMAN B S.

XX

PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;

XX WPI; 2004-061375/06.

DR

XX

XX New recombinant DNA construct comprising a promoter positioned to provide for expression of a polynucleotide encoding a polypeptide from a microbial source, useful for producing plants with improved properties.

PT

PS Claim 1; SEQ ID NO 7044; 122pp; English.

XX

XX The invention relates to a recombinant DNA construct comprising a promoter functional in a plant cell, where the promoter is positioned to provide for expression of a polynucleotide encoding a polypeptide from a microbial source. The invention also relates to a transformed plant comprising the recombinant DNA construct and a method of producing a transformed plant having an improved property. The plant is a crop plant such as maize or soybean. The method of producing a transformed plant having an improved property comprises transforming a plant with the recombinant DNA construct and growing the transformed plant, where the polynucleotide or polypeptide is useful for improving plant properties. The recombinant DNA construct is useful for producing plants with improved plant properties, e.g. improved cold, heat or drought tolerance, increased resistance to herbicides, extreme osmotic conditions, pathogens or pests, increased resistance to plant disease, better growth rate by modification of the cell cycle pathway with plant growth regulators, increased rate of homologous recombination, modified seed oil or protein yield and/or content, improved yield by modification of carbohydrate, nitrogen or phosphorus use and/or uptake, by modification of photosynthesis or by providing improved plant growth and development under at least one stress condition, improved lignin production or improved galactomannan production. This sequence represents a bacterial polypeptide used in the scope of the invention. Note: The sequence data for this patent did not form part of the printed specification but was obtained in electronic format from USPTO at seqdata.uspto.gov/sequence.html.

comprising the recombinant DNA construct and a method of producing a transformed plant having an improved property. The plant is a crop plant such as maize or soybean. The method of producing a transformed plant having an improved property comprises transforming a plant with the recombinant DNA construct and growing the transformed plant, where the polynucleotide or polypeptide is useful for improving plant properties. The recombinant DNA construct is useful for producing plants with improved plant properties, e.g. improved cold, heat or drought tolerance, tolerance to herbicides, extreme osmotic conditions, pathogens or pests, increased resistance to plant disease, better growth rate by modification of the cell cycle pathway with plant growth regulators, increased rate of homologous recombination, modified seed oil or protein yield and/or content, improved yield by modification of carbohydrate, nitrogen or phosphorus use and/or uptake, by modification of photosynthesis or by providing improved plant growth and development under at least one stress condition, improved lignin production or improved galactomannan production. This sequence represents a bacterial polypeptide used in the scope of the invention. Note: The sequence data for this patent did not form part of the printed specification but was obtained in electronic format from USPTO at seqdata.uspto.gov/sequence.html.

Sequence 502 AA;

Query Match 27.5%; Score 734.5; DB 8; Length 502;
Best Local Similarity 38.3%; Pred. No. 6.5e-54;
Matches 193; Conservative 73; Mismatches 217; Indels 21; Gaps 7;

Qy	19	HIDLDOAHAVASGGARIVLAPPARDRCRSEARLGAVIREARHVYGLTTGFGPLANRLIS	78
Db	10	HLTLAQLRRIAHFSETVOLCFSTFAAIDASAQAVENIVRKGPVYGVNTGFRLASTHIP	69
Qy	79	GENVRTLQANLVHHLASGVGVLDWTTTARAVLARIYSIAQASGASGEGTARIIDLINS	138
Db	70	LEQDLLQRLNLVSHAVGVEGPEPSRVVRLMIVLKVSLSLARGHSIRREVIALVTLINA	129
Qy	139	ELAPAVPSRGTVGASGDLTPLAHNVLCLOGRGDFLDRDGTGLDCAEGLRRGRLOPLDLSH	198
Db	130	DVLPVPIPKGSVGASGDLAPLAHMSAVLLGVGEVTIGE-RRASANEGLHAAGLEPFTLRA	188
Qy	199	RDALALVNGTSMTGIALVNAHACRHGNWAVALTALLAECLRGRTAWAAALSDLRPH	258
Db	189	KEGLALNGTQASTALALFNFAIEDLFTALVAGALSVDAAAASVVPFDARIHALRGHQ	248
Qy	259	GOKDAARLARVDGARVVRHVHTAERRLDAGDIGTEPEAGQDAYSLRCAPQVLGAGFDT	318
Db	249	GQDLSAAAYRALLEGSEINLSH-----SDCKVQDPYSLRCQPVWGCGLHQ	295
Qy	319	LAWHDRVLTIELNAVTDNPVFPDGSVPALHGGNFMGQHVALTSDALATATVTVLAGLAER	378
Db	296	MRHSAEVLCCANSVSDNPLIFPD-TCDLVSGGNFHAEPVAFADNLALAVAEIGALAE	354
Qy	379	QIARLTDERLNRLGPPFLHGRPGAGLNGFGMAQVATALLAEMRATG-PASIHISISTNAA	437
Db	355	RIALLIDTTLIS-GLPPEFLVR-DSGLNSGFMIAHVHTAAALASENKTLAHPASVDSLPTSAN	412
Qy	438	NQDVVSLGTIAARLCKEIKDRWAEITAILALCLAAAEELRCGSLDGVSPAGKKVQALR	497
Db	413	QEDHVSMTATFAARKLGDMASNTAHILAEILLAATQGDILR---EPHRTSMRLQSVTEAVR	469
Qy	498	EQFPFPLETDRPLGOEIAALATHLL	521
Db	470	RDVGHVHSDRHFSPDIAAAQVL	493

Search completed: December 4, 2005, 09:22:26
Job time : 190 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2005, 09:16:11 ; Search time 41 Seconds
(without alignments)
1234.391 Million cell updates/sec

Title: US-10-621-826-3

Perfect score: 2668

Sequence: 1 VKPMLMSPKPAVELDRHI.....RPLQGEIATLATHLLQSPV 526

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 80.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	728.5	27.3	508	1	UFBSSHs
2	712	26.7	511	2	F83897
3	711.5	26.7	524	2	F75610
4	708.5	26.6	511	2	E82228
5	706	26.5	516	1	JC1172
6	695	26.0	514	2	AG3040
7	695	26.0	514	2	E38245
8	694.5	26.0	657	2	S43415
9	688.5	25.8	511	2	H89758
10	679	25.4	677	2	T22333
11	677	25.4	509	2	C87368
12	670	25.1	511	2	G95945
13	666.5	25.0	657	2	A36087
14	662.5	24.8	657	2	A46128
15	657	24.6	506	2	A80596
16	652	24.4	509	2	G83009
17	643	24.1	510	1	A35251
18	620.5	23.3	510	2	B83009
19	615.5	23.1	510	2	AG0488
20	603.5	22.6	461	2	H84276
21	563	21.1	500	2	B98260
22	563	21.1	500	2	AG3024
23	554.5	20.8	710	2	S66313
24	543	20.4	705	2	JC5873
25	540	20.2	590	2	T09116
26	531.5	19.9	549	2	T05970
27	530	19.9	708	2	T14295
28	529	19.8	754	2	T09777
29	525	19.7	700	2	T06545

30	524	19.6	715	2	S66343	phenylalanine ammo
31	523.5	19.6	715	2	JQ2285	phenylalanine ammo
32	523	19.6	710	2	JC5872	phenylalanine ammo
33	519	19.5	725	2	S17444	phenylalanine ammo
34	516	19.3	718	2	S48726	phenylalanine ammo
35	515.5	19.3	716	2	S48725	phenylalanine ammo
36	515	19.3	712	2	S04127	phenylalanine ammo
37	514.5	19.3	716	2	S04463	phenylalanine ammo
38	514	19.3	720	2	S21174	phenylalanine ammo
39	511	19.2	713	2	S22991	phenylalanine ammo
40	510.5	19.1	667	2	T12749	phenylalanine ammo
41	509.5	19.1	701	2	S06475	phenylalanine ammo
42	509	19.1	712	2	T03663	phenylalanine ammo
43	507.5	19.0	716	2	S28185	phenylalanine ammo
44	506	19.0	723	2	S25303	phenylalanine ammo
45	505.5	18.9	721	2	A44133	phenylalanine ammo

ALIGNMENTS

RESULT 1

UFBSSHs

histidine ammonia-lyase (EC 4.3.1.3) huth [similarity] - Bacillus subtilis

N;Alternate names: histidase huth

C;Species: Bacillus subtilis

C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 09-Jul-2004

C;Accession: S18810; C69643; T47089

R;Oda, M.; Sugishita, A.; Furukawa, K.

J. Bacteriol. 170, 3199-3205, 1988

A;Title: Cloning and nucleotide sequences of histidase and regulatory genes in the Bacill

A;Reference number: S18808; MUID:88257040; PMID:2454913

A;Accession: S18810

A;Molecule type: DNA

A;Residues: 1-508 <ODA>

A;Cross-references: UNIPROT:P10944; UNIPARC:UPI00000471C6; EMBL:M20659; NID:g143074; PIDN

A;Experimental source: strain 1A270

R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero

C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi

A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.

Nature 390, 249-256, 1997

A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallerc

teich, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.;

Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,

A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel,

Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Pario, V.; Pohl, T.M.; Portetelle,

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,

A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror,

akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpetra, P.; Tognoni, A.; Tosato, V.; Uchiyama,

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.

A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

A;Reference number: A69580; MUID:98044033; PMID:9384377

A;Accession: C69643

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-508 <KUN>

A;Cross-references: UNIPARC:UPI00000471C6; GB:Z99124; GB:AL009126; NID:g2636442; PIDN:CAI

A;Experimental source: strain 168

R;Yoshida, K.; Sano, H.; Seki, S.; Oda, M.; Fujimura, M.; Fujita, Y.

Microbiology 141, 337-343, 1995

A;Title: Cloning and sequencing of a 29 kb region of the Bacillus subtilis genome contain

A;Reference number: Z24350; MUID:95219088; PMID:7704263

A;Accession: T47089

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-508 <YOS>

A;Cross-references: UNIPARC:UPI00000471C6; EMBL:D31856; NID:G603765; PIDN:BA06644.1; PII

A;Experimental source: strain BGSC1A1

C;Genetics:

A;Gene: huth

A;Function: catalyzes the formation of (E)-3-(1H-imidazol-4-yl)-propenoic acid (urocar

A:Pathway: histidine catabolism

C:Superfamily: Bacillus subtilis histidine ammonia-lyase

F;141-143/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted

F;142/Modified site: dehydroalanine (Ser) #status predicted

Query Match 27.3%; Score 728.5; DB 1; Length 508;

Best Local Similarity 35.5%; Pred. No. 1.5e-41;

Matches 182; Conservative 94; Mismatches 208; Indels 29; Gaps 9;

QY 20 IDLDQAHAVASGGARIVL-----APPARDRCASEARLVAVIREARHVYGLTTGGPPL 72

DB 2 VTLGGSLTTADVARVLDFFEEAAASEMERVKASRAVERIVDEKTIYINGTGFGEK 61

QY 73 ANRLISGENVRTLQANLVHHLASGVGLVDWTARTAMVLAARLVSIACGASGSEGTIARL 132

DB 62 SDVLIQKEDSAALQLNLILSHACGVGPFPECVSRAMLLLRANALLKGFSGVRAELIEQL 121

QY 133 IDLLNSLAPAVPSRGTVGASGDLTPLAHMVLCIQGRGDFLDRDGTLDGABGLRRGLQ 192

DB 122 LAFLNKRVHPVPIPOQSGSLGASGDLAPLSHLALALIGQGEVF-PEGERNPAMTGLKKAGIQ 180

QY 193 PLDLSHEDALVNGTSAMTGIALVNAHACHHLGNWAVALTALLAECICLRGTEAWAALS 252

DB 181 PVLTYSKEGALINGTOAMTAMGVAVIEAEKAYQTERIASLTIEGLQGIIDAFEDIH 240

QY 253 DLPHFPQKQDAARLARARVDGSAVRVHVIAERRLDAGDIGTPEAGQDAYSLRCAPOVL 312

DB 241 LARGYQOQIDVAERIRFVLSDSGLTTSQ--GELRV-----QDAYSLRCIPQVH 286

QY 313 GAGFDTLAHDRVLTIELNAVTDNVPFPPDGSPALHGNFMGQHVALTSDALATAVTVL 372

DB 287 GATWQTLGVYVKEKLEIEMNAATDNPLIFDNGD-KVISGNGFHGQPIAFAMDFLKIAISEL 345

QY 373 AGLAERQIARLTBERLNREGLPFLHRGPAGLNSGFMGAOVTATALLAEMRATG-PASISHS 431

DB 346 ANTAERIERELVNPQLN-DLPPFLSPHP-GLQSGAMTMQYAAASLVSENKTLAHPASVDS 403

QY 432 ISTNAANQDVVSLGTIARLCREKIDRWAEILALCLAAQAEELRCGSLDGVSAPGKK 491

DB 404 IPSANQEDHVSMTIARHAYQIVIANTRVIAIEALCALQAVEYR---GIEHAASYTKQ 460

QY 492 LVQALREQFPFLETRPLGQIEAALATHLQOS 524

DB 461 LFQEMRKVPSIQQDRVFSYDIERLTDWLKES 493

RESULT 2

F83897

histidase (histidine ammonia-lyase) huth [imported] - Bacillus halodurans (strain C 125)

C:Species: Bacillus halodurans

C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 09-Jul-2004

C:Accession: F83897

R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hira

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and

A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: F83897

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-511 <STO>

A:Cross-references: UNIPROT:Q9KBE6; UNIPARC:UPI000012CE89; GB:AP001513; GB:BA000004. NID

A:Experimental source: strain C-125

C:Genetics:

C:Superfamily: Bacillus subtilis histidine ammonia-lyase

Query Match 26.7%; Score 712; DB 2; Length 511;

Best Local Similarity 33.9%; Pred. No. 1.9e-40;

Matches 171; Conservative 102; Mismatches 209; Indels 22; Gaps 8;

QY 18 RHIDLDQAHAVASGGARIVLAPPARDRCASEARLVAVIREARHVYGLTTGGFPLANRLI 77

DB 10 RSLSLHDLHRIIYEGEETVSGASDESMEKVKQSRKAVQIIVADEKIIYGIITGPKFSDFI 69

QY 78 SGENVRTLQANLVHHLASGVGLVDWTARTAMVLAARLVSIACGASGASEGTIARLIDLNL 137

DB 70 DPDDVENLQHNLIYSHACGVGSPFPETVSRTMLVLRANALLKGFSGVRLPIVERLLALVN 129

QY 138 SELAPAVPSRGTVGASGDLTPLAHMVLCIQGRGDFLDRDGTLDGABGLRRGLRQLDLS 197

DB 130 ANIHPVPIPOQSGSLGASGDLAPLSHLALVLLGEGEVFK-GTKTKASFPALKEEIEPITLT 188

QY 198 HRDALALVNGTSAMTGIALVNAHACHHLGNWAVALTALLAECICLRGTEAWAALSDLRPH 257

DB 189 AKEGLALINGTOAMTAMGVAVIEAEKLAFAQSIIASLTWEGLRGIIIDAFDEQIFHARGY 248

QY 258 PGQKDAARLARARVDGSAVRVHVIAERRLDAGDIGTPEAGQDAYSLRCAPOVLGAGFD 317

DB 249 VEQVDVARRMESVLODSQITTRQ--GELRV-----QDAYSLRCIPQVHGATWQ 294

QY 318 TLAWHDRVLTIELNAVTDNVPFPPDGSPALHGNFMGQHVALTSDALATAVTVLAGLAE 377

DB 295 TLRVYVKEKLEIEMNAATDNPLIFDNGQ-KVISGNGFHGQIQLAMDFLGITAMAEALANISE 353

QY 378 RQIARLTBERLNREGLPFLHRGPAGLNSGFMGAOVTATALLAEMRATG-PASISHSISITNA 436

DB 354 RRTERLVNPQLN-DLPPFLSAAP-GVQSGVMTLQYCAASLVSENKTLAHPASVDSIPSSA 411

QY 437 ANQDVVSLGTIARLCREKIDRWAEILALCLAAQAEELRCGSLDGVSAPGKKLVQAL 496

DB 412 NQEDHVSMTIGISRHAYQIIONVRVLAIELICMVAQVDIR---GREKVASFTKKILEKG 468

QY 497 REQFPFLETRPLGQIEAALATHL 520

DB 469 REHVPIVDQDRMPAKDIERAAKWL 492

RESULT 3

F75610

histidine ammonia-lyase (EC 4.3.1.3) [similarity] - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans

C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004

C:Accession: F75610

R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; F

S.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma

S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A:Reference number: A75250; MUID:20036896; PMID:10567266

A:Accession: F75610

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-524 <WHI>

A:Cross-references: UNIPROT:Q9RZ06; UNIPARC:UPI000012CE8C; GB:AE001825; NID:

A:Experimental source: strain R1

C:Genetics:

A:Gene: DRA0147

A:Map position: 2

C:Superfamily: Bacillus subtilis histidine ammonia-lyase

C:Keywords: ammonia-lyase; carbon-nitrogen lyase

F;139-141/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted

F;140/Modified site: dehydroalanine (Ser) #status predicted

Query Match 26.7%; Score 711.5; DB 2; Length 524;

Best Local Similarity 36.8%; Pred. No. 2.1e-40;

Matches 188; Conservative 79; Mismatches 215; Indels 29; Gaps 9;

QY 16 LDRHDLQAHAVASGGARIVLAPPARDRCASEARLVAVIREARHVYGLTTGGFPLANR 75

DB 3 LDRDLNLEQFISVVRHGEQVELSAAARERARARTVIEQIVGDTPIYGVNTGFGKFENV 62

QY 76 LLSGENVRTLQANLVHHLASGVGLVDWTARTAMVLAARLVSIACGASGASEGTIARLIDL 135

DB 63 QIDRSOLAQHNLIIVSHAIGMEPLFAEVRVGMILLRAQSLSLGHSVGRVVEVVELLLAL 122


```

190 Db 190 EGLALLNGTQASTAFAL EGLFVAEDL FASATVCGAMSVEAALGSRPPDPDPRIHRVRGHRT 249
Qy 260 QKDAARLARARVDGSAFVRVHVIAERRLLDAGDIGTEPEAGQDAYSLRCAPQVLGAGFDTL 319
Db 250 QMDAATAYRHLLDVSSSEI-----QGSHNCEKVDQPFYSLRCQPOVMGACLOOI 297
Qy 320 AWHDRVLTIELNAVTDNP - VPPDGSVPALHGGNFMGOHVALTSDALATATVTVLAGLAER 378
Db 298 RSAAEVLVEVANSVSDNPLVFAEDGDI--ISGGNFHAEPVMAADNLALATAEIGLSER 355
Qy 379 QIARLTDERLNRGIPPLFHRGPAGINSGFMCAQVTTATALLAEMRATG-PASIHSTSTNNA 437
Db 356 RMLALLDSALSK-LPPFLVDN-GVNGSFMTAQVTTAAALASENKTTLAHPASVDSLPTSAN 413
Qy 438 NQDVVSLGTTAARLCREKIDRWASILATLALCLAQAAELRCGSGLDGVSYPAGKLVQALR 470
Db 414 QEDHVSWATTAARLRDNGENTRGILAVEYLAAAQGLDFRAPL----KSSPRIEEARQILR
Qy 498 EQFPFPLETDRPLGQEI AALATHLLQ 522
Db 471 EKVPFYDKRVFAPDIEK-ANALLQ 494

RESULT 5
JC1172
histidine ammonia-lyase (EC 4.3.1.3) [validated] - Streptomyces griseus
N:Alternate names: histidase

```

RESULT 5

JC1172
histidine ammonia-lyase (EC 4.3.1.3) [validated] - Streptomyces griseus
N/Alternate names: histidase
C/Species: Streptomyces griseus
C/Date: 14-May-1999 #sequence revision 14-May-1999 #text_change 09-Jul-2004
C/Accession: JC1172; PC1104; B42299
R/Wu, P.C.; Kroening, T.A.; White, P.J.; Kendrick, K.E.
Gene 115, 19-25, 1992
A/Title: Histidine ammonia-lyase from Streptomyces griseus.
A/Reference number: JC1172; MUID:92307427; PMID:1612436
A/Accession: JC1172
A/Status: nucleic acid sequence not shown
A/Molecule type: DNA
A/Residues: 1-516 <WUP>

A>Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.
 Nature 400, 477-483, 2000

A:Accession: PC1104
A:Molecule type: protein
A:Residues: 1-12 <WU2>
A:Cross-references: UNIPARC:UPI0000173007
R:Wu, P.C.; Kroening, T.A.; White, P.J.; Kendrick, K.E.
J. Bacteriol. 174, 1647-1655, 1992
A:Title: Purification of histidase from *Streptomyces griseus* and nucleotide sequence
A:Reference number: A4299; MUID:92165741; PMID:1537807
A:Accession: B42399
A:Molecule type: DNA
A:Residues: 1-516 <WUA>
A:Cross-references: UNIPARC:UPI00000471C7; GB:M77841; NID:g153316; PIDN:AAA2671
C:Genetics:
A:Gene: huth

A;Description: EC 4.3.1.3; histidine ammonia-lyase; catalyzes the formation of A;Pathway: histidine catabolism
A>Note: specific for L-histidine, shows no activity against D-histidine
A;Note: histidinal phosphate functions as competitive inhibitor; high concentrations , as in other histidases, dehydroalanine plays an important role in catalysis C;Superfamily: Bacillus subtilis histidine ammonia-lyase
C;Keywords: ammonia-lyase; carbon-nitrogen lyase
P;145-147/Cross-link: 5-imidazolinone (Cys-Gly) #status predicted
F;146/Modified site: dehydroalanine (Ser) #status predicted

Query Match 26.5% Score 706 DB 1: Length 516;
Best Local Similarity 39.6%; Pred No. 4.8e-40;
Matches 192; Conservative 212; Indels 24; Gaps 8

Qy 27 AVASGGARIIVLAPPARDCRASEARLGAVITREAHVYGLTTGFGPLANRLISGENVRYLQ 86
 ||||| :
Db 20 AVAHGHGARVELSAAVAEALAAARILVIDALAAKPEPVYGVSTGFAGALASRHGTTELRAQLQ 79

```
QY 87 ANLVHHLASGVPLDWTMTARVMVRLVSIQAQASGSEGTIARLIDLINSELAPAVPS 146
Db 80 RNIVRSHAAAGMPREVVRVRLMFLRLKTVASGHTGVRPEVAQTMADVLNAGITPVVHE 139
QY 147 RGTVGASGDLTPLAHMWLCLQGRGDFLDRDGTGLDGAEGLRGRGLQPLDLSHRDALVYN 206
Db 140 YGLSGCGSDDLAPLSHCHALTLMGGEAEGPGDGTVRPAGELLAHAHGIAPVELREKEGLALN 199
QY 207 GTSAMTGTIALVNAHACHRLGNVAVALTALLAECLLRGRTEAWAALSRLRPHPGQKDAAR 266
Db 200 GTDGMGLVMAALADLNLITSADITAALESALLGTDKVLAPELHAIHRPHPGQGSADN 259
QY 267 LRARVDSARVRVHVIARERLDAGDIGTEPEAGQDAYSLRCAPQVLGAGFDTTLAWHDRVL 326
Db 260 MSRLVAGSGLTGH-----QDDAPRV-----QDAYSVRCAPQVNGAGRDTLDHAALVA 307
QY 327 TIENAVTGNPNPFPDGSVPALHGGNFMGOHVALTSDALATATVTLVLAERQIARLTDE 386
Db 308 GRELASVDNPNVPLPDGRVES--NGNPHGAPVAVYVLDFLAIVAADLGSICERETDRLLDK 365
QY 387 RLNRGLPPFLHRGPAGLNSGFMGAQVATATALLAEM-RATGPASIHSTISNAQDVVSLG 445
Db 366 NRSHGLPPFL-ADDAGVDSGLMTAQYTOAALVSEMKEKLAVPASADSIIPSSAQEDHVS 424
QY 446 TIAARLCKRIDRWAETLAILCLQAABELRCGSLDGVSPAGKLVQALR----EQFP 501
Db 425 WSAARKLRATVDNLARIVAVELYEAATRAIELRAAEGLT-PAPASEAVVAALRAAGAEGPG 483
QY 502 PLETDRPLGQIEAALAT 518
Db 484 P---DRFLAPDLAADAAT 497

RESULT 6
AG3040
hiscidine ammonia-lyase huth [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
C:Accession: AG3040
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; W. O., I.
erge, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; Mclell
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-K. mm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AG3040
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-514 <KUR>
A:Cross-references: UNIPROT:Q8U827; UNIPARC:UPI000012CE88; GB:AE008689; PIDN:AAL447-1.1;
C:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: huth
A:Map position: linear chromosome
C:Superfamily: Bacillus subtilis histidine ammonia-lyase

Query Match 26.0%; Score 695; DB 2; Length 514;
Best Local Similarity 38.1%; Pred. No. 2.7e-39;
Matches 179; Conservative 66; Mismatches 203; Indels 22; Gaps 8;

QY 50 ARLGAVIREARHVYGLTTGFGPLANRLISGENVRTLQANLVHHLASGVGPLDWTMTARAM 109
Db 43 ARIAETAAGNAPVYGINTEGFKLASIKIDAADVATLQRNLILSHCCGVGVPLPENVRLLI 102
QY 110 VLARLVSIQAQASGSEGTIARLID-LLNSELAPAVPSRGTVCASGDLTPLAHMWLCLQG 168
Db 103 MALKVLSLGRGASGV-RLELVRLIEAMLEKGVIPVPEKGSVGSASGDLAPLAHMAAVMMG 161
QY 169 RGFDRDGTGLDGAEGLRGRLOPLDLSHRDALVNGTSAMTGTIALVNAHACHRLGNW 228
Db 162 EGEAF-YEGAPLPAGEALAKAGLTPVLAKEGLALINGTQTSTALAGLFAHRAAQA 220
QY 229 AVALTALLAECLLRGRTEAWAALSRLRPHPGQKDAARLARVDSARVRVHVIARERLD 288
Db 221 ALITGALSTDAAMGSSAPFPHDIHSLRGHKQIDAGAAALNLEGESEIRVSHIEGDERV- 279
QY 289 AGDIGTEPEAGDAYSLRCAPQVLGAGFDTTLAWHDRVLTIELNAVTDNPNPFPDGSVPAL 348
Db 280 -----QDPYCIRCPQVDGACLDLLRQVARTLIEANAVTDNPNLVLSDSNV--V 326
QY 349 HGGNFMGOHVALTSDALATATVTLVLAERQIARLTDELRNRLGPPFLHRGPAGLNSGFM 408
Db 327 SGNFHAEPVFAADQTLAICEIGAIAQRRVALLVDPALSYGLPAFLSKKP-GLNSGLM 385
QY 409 GAQVTTALLAEWRATG-PASTHSISTNAQDVVSLGTIAARLCKRIDRWAETLAILA 467
Db 408 EGEAF-YEGAPLPAGEALAKAGLTPVLAKEGLALINGTQTSTALAGLFAHRAAQA 220
```

Db 386 IAEVTSAAAMSENKQMAHPASVDSTPTTSANQEDHVSMAHGARRLLPMTDNLFAILGIEA 445

Qy 468 LCLAAQAEALRCGSLDGVSPAGKLVQALRQFPFPLETDRPLQGEIAALA 517

Db 446 LSAVQGVFLR---GFLKTSPELQKVIARLVVVVPSLEEDRYMAPDLKAAA 492

RESULT 8

S43415

histidine ammonia-lyase (EC 4.3.1.3) - human

N:Alternate names: histidase

C:Species: Homo sapiens (man)

C>Date: 07-Sep-1994 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004

C:Accession: S43415

R:Suchi, M.; Harada, N.; Wada, Y.; Takagi, Y.

Biochim. Biophys. Acta 1216, 293-295, 1993

A:Title: Molecular cloning of a cDNA encoding human histidase.

A:Reference number: S43415; MUID:94060103; PMID:7916645

A:Accession: S43415

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-657 <SUC>

A:Cross-references: UNIPROT:P42357; UNIPARC:UPI000012CE8E; DDBJ:D16626; NID:G451209; PID

C:Genetics:

A:Gene: GDB:HAL; HIS

A:Cross-references: GDB:120746; OMIM:235800

A:Map position: 12q22-12q23

C:Function:

A:Description: catalyzes the formation of (E)-3-(1H-imidazol-4-yl)-propenoic acid (iroce

A:Pathway: histidine catabolism

C:Superfamily: histidine ammonia-lyase

C:Keywords: ammonia-lyase; carbon-nitrogen lyase; histidine catabolism

F:253-255/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted

F:254/Modified site: dehydroalanine (Ser) #status predicted

Query Match 26.0%; Score 694.5; DB 2; Length 657;

Best Local Similarity 35.7%; Pred. No. 3.9e-39;

Matches 193; Conservative 81; Mismatches 227; Indels 39; Gaps 11;

Qy 6 AMSDFPSPQSGVYLYSKYREPEKYIELDGLRTTDLNVLGKRYKILTPTAERVQ 147

Db 88 AMSDFPSPQSGVYLYSKYREPEKYIELDGLRTTDLNVLGKRYKILTPTAERVQ 147

Qy 47 ASEARLGAIVREARHVYGLTGTGFPANRLNLSGENVTLQANLVHLSAGVGPVLDWTTA 106

Db 148 KSREVIDSIKEKTVVYGITGTGKFPARTVPIPKLQELQVNLVRSMSGVGPFLSPERC 207

Qy 107 RAMVLARLVSTAQASGASGSECTIARLDLNSLAPAPVPSRGTVGASGDLTPLAHMWLCL 166

Db 208 RMLLALRLNVLAKYSGISLETLKQVLEMFNASCLPYVPEKGTGASGDLAPLHSLALGL 267

Qy 167 QGRGDFLDRDGTDLRDGAELRRGRQLPDLIDLSHRDALALVNGTSAMTGTALVNAHACRHLG 226

Db 268 VGEKMWSPKSGWADAKVYLEHGLKPVILKPKGLALINGTQMTSLGCEAVERASAIA 327

Qy 227 NWAVALTALLAECLURGTEAWAALSDLRPHPGQKDAARLARVDSGARVRVHVIAERR 286

Db 328 ROADIVAALTLEVLKGTTKAPDTHALRPHRGQIEVAFRPSLLDSDHHPSEIAESHRF 387

Qy 287 LDAGDIGTEPAGODAYSRLCAPQVLGAGPTTLAHHDRVLTIELNATDNP-VFPPPGSV 345

Db 388 CD-----RVQDAYTLRCCPQVGHVNDTIAFVKNIIITTELSATDNPVFANRGE- 437

Qy 346 PALHGGNFMGQHVALTSDALATATVTLAGLAERQIARLTDERLNRLGPPPLHRRGPAGLNS 405

Db 438 -TVSGGNFHEGYPKALDYLAIGHLELAISERIERLCNPSLSE-LPAFL-VAEGGLNS 494

Qy 406 GFMGQAVTATALLAEMRA-TGPASIHSTNAANQDVVSLGTITAAIRCEREKIDRWAEILA 464

Db 495 GFMIHCTAAALVSENKALCHPSSVDSLSTSAATEDHVSNGGAARALRVIEHVEQVLA 554

Qy 465 ILALCLAAQAE-LRCGSLDGVSPAGKLVQALRQFPFPLETDRPLQGEIAALATHLQ 523

Db 555 IELLAACQIEFLR---PLKTTTTL-ELKYVDLYRSVVVRPWIKDRPFMAPDIEAAHRLLEQ 610

RESULT 9

H89758

histidine ammonia-lyase [imported] - Staphylococcus aureus (strain N315)

C:Species: Staphylococcus aureus

C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001

C:Accession: H89758

R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguchi, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; Shibata, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.

Lancet 357, 1225-1240, 2001

A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.

A:Reference number: H89758; MUID:21311952; PMID:11418146

A:Accession: H89758

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-511 <KUR>

A:Cross-references: UNIPARC:UPI0000165A2C; GB:BA000018; PID:G13699925; PIDN:BA041224.1; C

C:Genetics:

A:Gene: hnhH

C:Superfamily: Bacillus subtilis histidine ammonia-lyase

Query Match 25.8%; Score 688.5; DB 2; Length 511;

Best Local Similarity 35.4%; Pred. No. 7.2e-39;

Matches 174; Conservative 89; Mismatches 208; Indels 21; Gaps 7;

Qy 33 ARIVLAPPDRDCRSEARLGAIVREARHVYGLTGTGFPANRLNLSGENVTLQANLVH 92

Db 30 SKIEIIDDALEVRKKSRAVVERIENETVYGITGTGFLSFDVRIDPTQNELVNLI 89

Qy 93 LASGVGPVLDWTTARAMVLARLVSTAQASGASGSECTIARLDLNSLAPAPVPSRGTVGA 152

Db 90 HACGLGEPFSEKVALVMMILRLNTLLKGSGATLELVRLQOFFINERIIPIPOGSLGA 149

Qy 153 SGLDITPLAHMWLCLQGRGDFLDRDGTDLRDGAELRRGRQLPDLIDLSHRDALALVNGTSAMT 212

Db 150 SGLDITPLAHMWLCLQGRGDFLDRDGTDLRDGAELRRGRQLPDLIDLSHRDALALVNGTSAMT 208

Qy 213 GIALVNAHACRHLGNNAVALTALLAECLRGRTAWAALSRLRPHPGQKDAARLARVD 272

Db 209 AQGVSYIEAEIDLYQSEWIAALTHQSLNGIIRYHSDHVSVRNFQEQINVAARMRWLE 268

Qy 273 GSARVVRHVIAERLRDAGDIGTEPAGQDAYSRLCAPQVLGAGPTTLAHHDRVLTIELN 332

Db 269 GSTLTTRQ--AEIRV-----QDAYTLRCIPQIHGASQVFNYYKQLEFENNA 314

Qy 333 VTQNP-VFPPPGSVPAHLHGNFMGQHVALTSDALATATVTLAGLAERQIARLTDERLN 391

Db 315 ANDNPLIFEEANETFTVSGGNFHQPTAFALDHLKLGSELANVSERRLERLVNPQLNGD 374

Qy 392 LPDPLHRRGPAGLNSGFMAQVTTATALLAEMRATG-PASIHSTNAANQDVVSLGTITAA 450

Db 375 LPAPLSPEP-GLOSGAMIMQYAAASLVSENKTLAHPASVDSITSSANQEDHVSMTGAAR 433

Qy 451 LCEREKIDRWAEILAILALCLAAQAEALRCGSLDGVSPAGKLVQALRQFPFPLETDRPLG 510

Db 434 HGQIENARRVLAIECVIALQAAELK--GVGSLSPKTRKRYEEFVSIVPSITHDRQFH 490

Qy 511 QETAAALATHLLQ 522

Db 491 KDIEAQAQYLKQ 502

RESULT 10

T22333

histidine ammonia-lyase (EC 4.3.1.3) - Caenorhabditis elegans

N:Alternate names: histidase; protein F47B10.2

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C:Accession: T22333

R;Harris, B.
submitted to the EMBL Data Library, November 1995
A;Reference number: Z19550
A;Accession: T22333
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-677 <WIL>
A;Cross-references: UNIPROT:Q20502; UNIPARC:UPI000012CE8A; EMBL:Z68004; PIDN:CAA919/2.1;
A;Experimental source: clone F47B10
C;Genetics:
A;Gene: CESP:F47B10.2
A;Map position: X
A;Introns: 24/3; 77/1; 111/2; 178/1; 255/1; 335/1; 395/1; 523/1; 544/3; 604/2
C;Superfamily: histidine ammonia-lyase
C;Keywords: ammonia-lyase; carbon-nitrogen lyase
F;269-271/Cross-link: 5-imidazolone (Cys-Gly) #status predicted
F;270/Modified site: dehydroalanine (Ser) #status predicted

Query Match 25.4%; Score 679; DB 2; Length 677;
Best Local Similarity 35.3%; Pred. No. 4.5e-38;
Matches 182; Conservative 85; Mismatches 231; Indels 18; Gaps 10;

QY 3 PMLAMSPKPAVELDRHIDL-DQAHAVASGARIVLAPPARDRCRASEARLGAVIREARH 61
DB 120 PMV-LAPPTKLLILDGNLLPEDLVCKEKGECQAIQLSMESEDRIKARTFLEKIASEHRA 178

QY 62 VYGLTTGFGFLANRLISGENVRTLOANLVHHLASGVGPVLDWTTARAWLARLVSTAOGA 121
DB 179 VYGVTTGFGFTSNVTPPEKKLQLNLIRSHATGGEPLAPNARWMLALRINIILAKGH 238

QY 122 SGASEGTIARLIDLLINSELAPVPSRGTVGASGDLTPLAHMVLCLQGRGDFLDRDGTRLD 181
DB 239 SGISVENIKKMAAFNAFCVSVYVQQGTGCGSDGLCPLAHLALGLLGEGKMWSPSTTGWQP 298

QY 182 GAGLGRGRLOPLDLSRDALALVNGTSAMTALVNNAHCRHLGNWALVALLAECLR 241
DB 299 ADVLKNKLEPLELPGKEGLALINGTQMTALGAYTLERAHNIARQADVAIALSLDLVK 358

QY 242 GRTEAWAALSDLRPHGQKAAARLRARVDGSRVVRHVIAERLRDAGDIGTEPEAGOD 301
DB 359 GTTRAVDPDHRTPRHGQNLRLALLHSEANSQ--IASHRNCTKV-----QD 409

QY 302 AYSLCAPQVLGAGFTTLAWHDRLVLTIELNAVDNPNFPPDGSPALHGNFMGQHVALT 361
DB 410 AYTLRCVPQVGVVHDTIEPVREIITMNSATDNPLVPADRE-EIISGNGPHGEYPAKA 468

QY 362 SDALATAVTVLAGLAERQIARLDRLNRCLPPLPHRGAGLNSGFMGQVTTATALLAEM 421
DB 469 LDFLATAVAELAQMSERRERLVNKELS-GLPTFL-TPDGGLNSGFMVTQLCAASLVSEN 526

QY 422 RA-TGPASIHSTNAANQDVSLGTIAARLCREKIDRWAEIILAILALCLAAQAEELRCGS 480
DB 527 KVLCHSSVDVSIPTSCNQEDHVSNGFPAARKALTVVEHVAVLMBELLAAQCQIEFL--K 584

QY 481 GLDGVPSPAGKKLVQALREQPPPLETDRPLGQETAAAL 516
DB 585 PLISTAPL-HKIQLVRSVAPPLNEDRYMKPEIDAV 619

RESULT 11
C87368
histidine ammonia-lyase [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Jul-2004
C;Accession: C87368
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Jol
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: C87368
A;Status: preliminary

A;Molecule type: DNA
A;Residues: 1-509 <STO>
A;Cross-references: UNIPROT:P58082; UNIPARC:UPI000012CE8B; GB:AE005673; NID:gl3422237; PJ
C;Genetics:
A;Gene: CC0959
C;Superfamily: Bacillus subtilis histidine ammonia-lyase

Query Match 25.4%; Score 677; DB 2; Length 509;
Best Local Similarity 37.9%; Pred. No. 4.3e-38;
Matches 187; Conservative 61; Mismatches 215; Indels 30; Gaps 9;

QY 32 GARIVLAPPARDRCRASEARLGAVIREARHVYGLTTGFGFLANRLISGENVRTLOANLVH 91
DB 22 GASARLAESAWPVIABSAVAQVIRILAKGEPVYGVNTGFGKLASVRIGDADLETQRNIVL 81

QY 92 HLASGVGPVLDWTTARAWLARLVSTAOGASGSECTIARLIDLLINSELAPVPSRGTVG 151
DB 82 SHAAAGVGPEPVPVIRLMMALKLASLAQAQSGVRVETVRMLEEMVLEGLTPVVPVPCQGSVG 141

QY 152 ASGDLTPLAHMVLCLQGRGDFLDRDGTRLDGAELRRGRLOPLDLSHRDALALVNGTSAM 211
DB 142 ASGDLAPLSHMAATMIGVGEIF-VGQRLPAAQAQAAGLEPLTLGPKGGLALLNGTQPS 200

QY 212 TGIALVNNAHCRHLGNWALVALLAECLRGRTEAWAALSDLRPHGQKDAARLRARV 271
DB 201 TANALAGLFEAERLFQSAVLVTGALSTEAAGSGDTPDPRIHTLRRHVQGIETAAALRALM 260

QY 272 DGSARVVRHVIAERLRDAGDIGTEPEAGODAYSLCAPQVLGAGFTTLAWHDRLVLTIELN 331
DB 261 SASEIRASHL-----KEDERVQDPFYCLRCQCPQVMGAALDILRQAATTTLATEAN 308

QY 332 AVTDNPNFPPDGSPALHGNFMGQHVALTSDALATAVTVLAGLAERQIARLTDERLNRG 391
DB 309 CVSDNPLIFPEAD-EALSGNFHAPVAFADMIALAVCEIGSIAERRIAMLVDPALSG 366

QY 392 LPPFLHRCGAGLNSGFMGQVTTATALLAEMRATG-PASIHSTISSTNAANQDVVSLGTIAAR 450
DB 367 LPAFLTPKP-GLNSGFMIPQVTAALVSENKQRAVPASVDSIPTSANQEDHVSMAAHGAR 425

QY 451 LCREEKIDRWAEIILAILALCLAAQ-----AELRCGSLDGVPSPAGKKLVQAL-REQPPLET 505
DB 426 RLLAVVENADAVLGIELLAAAGCDFHAPLRSAALEA-----VRLATRSKVPHLSD 477

QY 506 DRPLGQETAAALAT 518
DB 478 DRHFHPDMEAAANT 490

RESULT 12
G95945
probable histidine ammonia-lyase (EC 4.3.1.3) [imported] - Sinorhizobium meliloti (strain
C;Species: Sinorhizobium meliloti
C;Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 09-Jul-2004
C;Accession: G95945
R;Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernanc
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A;Title: The complete sequence of the 1.683-kb pSymB megaplasmid from the N2-fixing endo
A;Reference number: A95842; MUID:21396508; PMID:11481431
A;Accession: G95945
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-511 <KUR>
A;Cross-references: UNIPROT:O31197; UNIPARC:UPI000012CE93; GB:AL591985; PIDN:CAC49231.1;
A;Experimental source: strain 1021, megaplasmid pSymB
R;Gallbert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
hebault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A;Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A;Reference number: A96039; MUID:21368234; PMID:11474104
A;Contents: annotation

Db 170 FGKPARTVTPANKLQELQVNLVVRSHSSGVGKPLSPERCMLLALRINVLAKYVSGISLET 259
 Qy 129 IARLIDLLNSELAPAVPSRGTVGASGDLTPLAHVMVLCIQGRGDFLDROGTRLDGAEGRLR 188
 Db 230 LKQVIEVFNASCLSYVPEKGTGVGASGDLAPLSHLALGLIGEGKMWSPKSGWADAKYVLEA 289
 Qy 189 GRLOPLDLSHRDALALVNGTSSAMTGIALVNVNAHAACHLGNWAVALTALLAECLRGRTAWA 248
 Db 290 HGLKPIVLKPKEGIALINGTQWITSLGCEAVERASAIARQADIVAALTLEVLKGTTKAFD 349
 Qy 249 AALSDLRPHPCQDKAARLARVARVDSGARVVRHVTAERRLDAGDICTEPEAGODAYSIRCA 308
 Db 350 TDHAVRPHGQIEVAFRFRSLSDSDHHHSIABSHRFCD-----RVQDAYTLRCC 400
 Qy 309 PQVLGAGDFDTLAWHDRVLTITELNAVTDNP-VFPDPGSPVAPLHGNGFMGOHVALTSDALAT 367
 Db 401 PQVHGVVNDITAFVKDIITTELNSATDNPVPSRGE--TISGGNFHGEYPAKALDYLA 458
 Qy 368 AVTVLAGIARQIARLTDRLNRGLPPPLHRRGPAGLNSFGMGAAQVATATALLAEMRA-TGP 426
 Db 459 GVHLEAAISERRIERLCNPSLSE-LPAFL-VAEGGLNSGFMIAHCTAAALVSESALCHP 516
 Qy 427 ASHSISITNAANDVSLGTITTAARLCREKIDRWAEIILAILALCLAAQAB-IRCGSLDGV 485
 Db 517 SSVDLSLTSAAEDHVHSGMGWAARKALRVIEHVEQVLAIELLAACQGIETFLR---PLKTT 573
 Qy 486 SPAGKLVQALREQFPFPLETRPLQGEIATAALATHLQQ 523
 Db 574 TPL-EKYVDLVRVVRPWIKDRFMAPDIEAHRLLDQ 610

RESULT 14
 A46128
 histidine ammonia-lyase (EC 4.3.1.3) - mouse
 N:Alternate names: histidase
 C:Species: Mus musculus (house mouse)
 C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
 C:Accession: A46128
 R:Taylor, R.G.; Grieco, D.; Clarke, G.A.; McInnes, R.R.; Taylor, B.A.
 Genomics 16, 231-240, 1993
 A:Title: Identification of the mutation in murine histidinemia (his) and genet
 A:Reference number: A46128; MUID:93252384; PMID:8486363
 A:Accession: A46128
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-657 <RAY>
 A:Cross-references: UNIPROT:P35492; UNIPARC:UPI0000003F1B; GB:I07645; NID:G193
 A:Experimental source: C57BL
 A:Note: sequence extracted from NCBI backbone (NCBIN:131641, NCBI:P:131644)
 C:Function:
 A:Description: catalyzes the formation of (E)-3-(1H-imidazol-4-yl)-propenoic a
 A:Pathway: histidine catabolism
 C:Superfamily: histidine ammonia-lyase
 C:Keywords: ammonia-lyase; carbon-nitrogen lyase; histidine catabolism
 F:253-255/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted
 F:254/Modified site: dehydroalanine (Ser) #status predicted

Query Match 24.8%; Score 662.5; DB 2; Length 657;
 Best Local Similarity 35.5%; Pred. No. 5.5e-37;
 Matches 184; Conservative 81; Mismatches 232; Indels 21; Gaps 10

Qy 10 PKPAVELD-RHIDLDQAHAVASGGARIVTLAPPARDRCRASEARLGAIVREARHVYGLTGG 68
 Db 110 PEKYIALDGSLSLTDLVNLGKRYKILKLSIAEKVQOSREVIDSIKERTVTVYITGG 169
 Qy 69 FGPLANRLISGENVRTLQANLVHHLASGVGVPLDWTMTTARAMVTLARLVSAIQAGASBGT 128
 Db 170 FGKPARTVTPANKLQELQVNLVVRSHSSGVGKPLSPERCMLLALRINVLAKYVSGISLET 229
 Qy 129 IARLIDLLNSELAPAVPSRGTVGASGDLTPLAHVMVLCIQGRGDFLDROGTRLDGAEGRLR 188
 Db 230 LKQVIEAFNASCLSYVPEKGTGVGASGDLAPLSHLALGLIGEGKMWSPKSGWADAKYVLEA 289

QY 189 GRLOPLDLSHRDALALVNGTSAMTGIALVNAHACRHLGNWAVALTALLAECLRGTRTEAWA 248
Db 290 HGLKPIVLKPEGLALINGTOMITSLGCEALERASAIARQADIVAALTLEVLKGTTKAPD 349
QY 249 AALSDLRPHGQKDAARLARARVDGSGARVVRHVIAERLDAGDIGTEPEAGQDAYSLRCA 308
Db 350 TDHAVRPHRGQTEVAFRFRSLDSDHPHSEIAESHRFC-----RVQDAYTLRCC 400
QY 309 PQVLGAGFTLAWHDRVLTITLNAVTNPN-VFPDGSVPALHGGNFMGOHVALTSDALAT 367
Db 401 PQHGVVNDTIAFVKDIITELNSATDNPMVFASRGE--TISGNFHFGEYPAKALDYLA 458
QY 368 AVTVLAGLAERQIARLTDERLNRGLPPFLHRGPAGLNSGFMGAQVTTATALLAEMRA-TGP 426
Db 459 GVHELAAISERRIERLCNPSLSE-LPAFL-VASGGLNSGFMIAHCTAAALVSESKALCHP 516
QY 427 ASHISISTNANQDVSLGTIARLCREKIDRWAEIILAILALCLAQAAE-LRCGSGLDGV 485
Db 517 SSVDSLSTSAATEDHVSIMGWAARKALRVVHVVEQVLATIELLAACQIGBFLR---PLKTT 573
QY 486 SPAGKKLVQALREQFPFPLETDRPLGQEIATLALHLLQQ 523
Db 574 TPL-EKVDLVRSVVRPWIKDRFMAPDIEAAHRLLDQ 610
RESULT 15
AE0596
histidine ammonia-lyase (EC 4.3.1.3) [imported] - Salmonella enterica subsp. enterica serovar typhi
C:Species: Salmonella enterica subsp. enterica serovar typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AE0596
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Garra, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar typhi
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AE0596
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-506 <PAR>
A:Cross-references: UNIPARC:UPI00005A0F7; GB:AL513382; PIDN:CAD05239.1; PID:gl6502.09;
C:Genetics:
A:Gene: hutH
C:Superfamily: Bacillus subtilis histidine ammonia-lyase
C:Keywords: ammonia-lyase; carbon-nitrogen lyase
Query Match 24.6%; Score 657; DB 2; Length 506;
Best Local Similarity 36.4%; Pred. No. 9.4e-37;
Matches 186; Conservative 76; Mismatches 215; Indels 34; Gaps 11;
QY 20 IDLDQAHAVASGARIVLAPPARDRCRSEARIGAVIREARHYGLTTGFGPLANRLISG 79
Db 11 LSLQLYDVRHVPQRLDASALDGINASVACVNDIVAEGRYAGINTGFGLLAQTRIAD 70
QY 80 ENVRTLOANLVHLSGCVGLDWTTPARAVLARLVSIAGGASGASEGTIARLIDLINSE 139
Db 71 EDLQNLQSLVLSHAAGVGDPLDDAMVRLIMVLKINSLARGFSGIRLSVIEALIALVNA 130
QY 140 LAPAVPSRGTVGASGDLTPLAHMVLCLQGRGDFLDRDGTDLDCAGELRRGRQLPLDLSHR 199
Db 131 VYPLIPAKGSGVAGDLAPLAHLSTLLGEGK-ARWQEWLPAQAALKKAGLEPVVALAK 189
QY 200 DALALVNGTSAMTGIALVNAHACRHLGNWAVALTALLAECLRGTRTEAWAALSDLRPHFG 259
Db 190 EGLALINGTOASTAFALRGLFEAQELFASAVCGALTTEAVLGSRCRPFDIRIHAARGQQG 249
QY 260 OKDAAARLARVDGSGARVVRHV-----IAERLDAGDIGTEPEAGQDAYSLRCAPOVLG 313
Db 250 QIDV-----ARLFRHLTDTSAIESHHCHKV-----QDPYSLRCQPQVNG 291

QY 314 AGFDTLAWHDRVLTITELNAVTNPN-VFPDGSVPALHGGNFMGOHVALTSDALATAVTVL 372
Db 292 ACITQLRQTKKEVLLAEANAVSDNPLVFAAGEV--ISGNGFHAEPVMAADNLALAI AEI 349
QY 373 AGLAERQIARLTDERLNRGLPPFLHRGPAGLNSGFMGAQVTTATALLAEMRATG-PASIIHS 431
Db 350 GALSERRITLMDKQMSQ-LPPFLVKN-GGVNSGFMIAQVTTAAALASENKALAHPHSVDS 407
QY 432 ISTNANQDVSLGTIARLCREKIDRWAEIILAILALCLAQAAE-LRCGSGLDGVSPAGKK 491
Db 408 LPTSANQEDHVSMAAPAGRRLWEMAANTRGVIAVEWLAACQIGIDLR--EGLTS-SPLEEQ 464
QY 492 LVQALREQFPFPLETDRPLGQEIATLALHLLQQ 522
Db 465 ARQTLERVAHYTQDRFFAPDIECATITLLAQ 495

Search completed: December 4, 2005, 09:27:09
Job time : 43 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2005, 09:15:56 ; Search time 232 Seconds
(without alignments)
1599.603 Million cell updates/sec

Title: US-10-621-826-3

Perfect score: 2668

Sequence: 1 VKPMLAMSPRPVAVELDRHI.....RPLQGEIAALATHLLQQSPV 526

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	851	31.9	507	1 HUTH_SYMTH	Q67jh4 symbiobacte
2	814.5	30.5	539	2 Q8GMG0 STRGL	Q8gmgo streptomyc
3	814	30.5	508	2 Q4NR19 9DEL	Q4nr19 anaeromyxob
4	801	30.0	508	1 HUTH_THETN	Q8rbh4 thermoaer
5	797.5	29.9	514	1 HUTH_GLOVI	Q7ncb3 gloeobacter
6	792.5	28.7	523	2 Q9KHJ9 9ACTO	Q9khj9 streptomyc
7	763	28.6	499	2 Q4H6E7 9DRIO	Q4he67 deinococcus
8	762.5	28.6	526	1 HUTH_AZOSE	Q5nrx8 azoarcus sp
9	753	28.2	510	1 HUTH_STIAU	Q93tx3 stigmatella
10	749.5	28.1	507	1 HUTH_BURVA	Q62lj6 burkholderi
11	749.5	28.1	507	1 HUTH_BURPS	Q63eh6 burkholderi
12	746	28.0	507	1 HUTH_TREDE	Q73q56 treponema d
13	733	27.5	504	1 HUTH_GEOKA	Q51310 geobacillus
14	729.5	27.3	507	2 Q4LS33 9BURK	Q4ls33 burkholderi
15	728.5	27.3	508	1 HUTH_BACSU	P10944 bacillus su
16	726.5	27.2	519	1 HUTH_BRAJA	Q89gv3 bradyrhizob
17	726	27.2	514	1 HUTH_CLOTE	Q891q1 clostridium
18	726	27.2	524	2 Q5Z0G8 NOCPA	Q5z0g8 nocardia fa
19	726	27.2	528	2 Q8XSV2 RALSO	Q8xsv2 ralstonia s
20	725	27.1	511	1 HUTH2_FUSNN	Q8rdu4 fusobacteri
21	723.5	27.1	516	2 Q7PSN4 FUSNV	Q7psn4 streptomyc
22	721.5	27.0	512	1 HUTH_STRCO	Q9swl1 streptomyc
23	721	27.0	500	1 HUTH_BACSK	Q5waz6 bacillus cl
24	715	26.8	495	1 HUTH_THEVO	Q978n8 thermoplas
25	715	26.8	512	1 HUTH_STRAW	Q82133 streptomyc
26	712	26.7	511	1 HUTH_BACHD	Q9kbe6 bacillus ha
27	711.5	26.7	524	1 HUTH_DESPS	Q6akp3 desulfotale
28	711.5	26.7	524	1 HUTH_DEIRA	Q9rz06 deinococcus
29	710.5	26.6	541	2 Q84FL5 ENTAG	Q84fl5 enterobacte
30	708.5	26.6	510	1 HUTH_CHRVO	Q7p188 chromobacte
31	708.5	26.6	511	1 HUTH_VIBCH	Q9ksq4 vibrio chol

32 706 26.5 496 1 HUTH_THRAC Q9hli6 thermoplasm
33 706 26.5 514 1 HUTH_STRGR P24221 streptomyc
34 701.5 26.3 513 1 HUTH_SHEON Q8ekj4 shewanella
35 700 26.2 511 1 HUTH_VIBPA Q87q77 vibrio para
36 699.5 26.2 515 1 HUTH_GLUOX Q5frr8 gluconobact
37 698.5 26.2 510 1 HUTH_VIBF1 Q5e0c6 vibrio fise
38 696 26.1 925 1 HUTHI_BRUME Q8ydo9 brucella me
39 695.5 26.1 657 2 Q4VB93 HUMAN Q4vb93 homo sapien
40 695 26.0 514 1 HUTH_AGRY5 Q8u8z7 agrobacteri
41 694.5 26.0 657 1 HUTH_HUMAN P42357 homo sapien
42 694.5 26.0 657 2 Q4VB92 HUMAN Q4vb92 homo sapien
43 692.5 26.0 509 2 Q4IY51 AZOVI Q4iyl1 azotobacter
44 689 25.8 664 2 Q6GMC2 XENLA Q6gmc2 xenopus lae
45 688.5 25.8 504 1 HUTH_STAMM P64415 staphylococ

ALIGNMENTS

RESULT 1

HUTH_SYMTH
ID HUTH_SYMTH STANDARD; PRT; 507 AA.
AC Q67JH4;
DT 13-SEP-2005 (Rel. 48, Created)
DT 13-SEP-2005 (Rel. 48, Last sequence update)
DE Histidine ammonia-lyase (EC 4.3.1.3) (Histidase).
GN Name=hutH; OrderedLocusNames=STH3194;
OS Symbiobacterium thermophilum.
OC Bacteria; Actinobacteria; Symbiobacterium.
OX NCBI_TaxID=2734;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=T / IAM 14863;
RX PubMed=15383646; DOI=10.1093/nar/gkh830;
RA Ueda K., Yamashita A., Ishikawa J., Shimada M., Watsuji T.,
RA Morimura K., Ikeda H., Hattori M., Beppu T.;
RT "Genome sequence of Symbiobacterium thermophilum, an uncultivable
bacterium that depends on microbial commensalism.";
RL Nucleic Acids Res. 32:4937-4944 (2004).
CC -!- CATALYTIC ACTIVITY: L-histidine = urocanate + NH(3).
CC -!- PATHWAY: Histidine degradation; first step.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -!- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
which is formed autocatalytically by cyclization and dehydration
of residues Ala-Ser-Gly (By similarity).
CC -!- SIMILARITY: Belongs to the PAL/histidase family.
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use as long as its content is in no way modified and this statement is not
removed.

CC EMBL; AP006840; BAD42176.1; -, Genomic_DNA.
DR HAMAP; MF_00229; -, 1.
DR InterPro; IPR005921; HutH.
DR InterPro; IPR001106; Phe/His_NH3lyase.
DR InterPro; IPR005829; Sug_transporter.
DR Pfam; PF00221; PAL; 1.
DR TIGRFAMs; TIGR01225; hutH; 1.
DR PROSITE; PS00488; PAL_HISTIDASE; 1.
KW Complete proteome; Histidine metabolism; Lyase.
FT MOD_RES 143 143 2,3-didehydroalanine (Ser) (By
similarity).
FT CROSSLINK 142 144 5-imidazolinone (Ala-Gly) (By
similarity).
SQ SEQUENCE 507 AA; 53709 MW; 4343495E574F09B5 CRC64;

Query Match 31.9%; Score 851; DB 1; Length 507;

Best Local Similarity 42.8%; Pred No. 2,1e-45;

Matches 217; Conservative 69; Mismatches 197; Indels 24; Gaps 9;

```
Qy 13 AVELDRHIDLOAHAVAGGARIIVLAPPARDCRASEARLGAIVTREAHHVYGLTGTGGL 72
Db 3 AVELGAHLTPEVVARHGARVLTTEVQRVARASEMVERLVRPPVGIITGFKF 62
Qy 73 ANRLISENVRTLOANLVHSLASGVPLDWTTRAMVLAELVLSIAQASGASGTTARL 132
Db 63 SDVPISAEQTAQRNLMHSHACAVGBPLAAEVVRAMLLLRQAQLSRHSGIRAEITL 122
Qy 133 IDLNSSELAPVPSRGTVGASGDLTPLAHMVLCLQGRGDFDRDGTDLDAEGLRRGRK 192
Db 123 VAFNLGLTVPVPGSGISGASGDLAPLAHMSLPLIGLGEAV-VNGERLSGAEALQVR 181
Qy 193 PLDSLHSDALALVNGTSAMTGIALVNAHACHLGNWAVALTALLAECLRGTEAWAALS 252
Db 182 PLTLTAKEGLALINGTOAMTALGSLGLHDAQVLLKTADIAAAMTAEGALPAAMDPR 241
Qy 253 DLRPHGKQDAARLARVDSARVVRHVAERRLDAGDIGTEPEAGQDAYSRLCAPVL 312
Db 242 ALRLHTQQAARNRLRLTEGSRLLTTR-----PGQWRT-----ODPYTLRCLP 287
Qy 313 GAGFDTLAWHDRVLTIELNAVTDNP-VPPPDGSPALHGGNFMGQHVALTSDALATAV 371
Db 288 GASRTAEHVAQVLDWENNAVTDNPLFPDDDEV--ISGGNFHGQPVALLDYLAI 345
Qy 372 LAGLAERQIARLTDERLNRLGPPFLHRGPAGLNSGFMGAQVATALLAEWRATG-P 430
Db 346 LGDIAERRIERNLPOLS-GLPAFLTRN-GGVHSGMLTITVTAASLVSEKNVLAHP 403
Qy 431 SISITNAQDVVSLGTIAARLCREKIDRWAEIILAILALCAQAEELRCGSLDGVSP 490
Db 404 SIPSSANOEDHVSMTGAARARQVIANRRVLAIELLCAQAQLEF---VGPERLA 460
Qy 491 KLVALREQFPPLTDRPLGOEIAALA 517
Db 461 AYAARIRVAPLSDRVAPADIEALA 487
RESULT 2
Q8GMG0 STRGL PRELIMINARY; PRT; 539 AA.
AC Q8GMG0;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Putative ammonia lyase/transferase.
OS Streptomyces globisporus.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1908;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C-1027;
RX MEDLINE=22171413; PubMed=12183628; DOI=10.1126/science.1072110;
RA Liu W., Christenson S.D., Standage S., Shen B.;
RT "Biosynthesis of the enediyne antitumor antibiotic C-1027.";
RL Science 297:1170-1173(2002).
DR EMBL; AY048670; AAL06680.1; -; Genomic_DNA.
DR HSSP; P21310; 1GKM.
DR GO; GO:0005737; Cytoplasm; IEA.
DR GO; GO:0016211; F:ammonia lyase activity; IEA.
DR GO; GO:0004397; F:histidine ammonia-lyase activity; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0009058; P:biosynthesis; IEA.
DR GO; GO:0006548; P:histidine catabolism; IEA.
DR InterPro; IPR005921; Huth.
DR InterPro; IPR001106; Phe/His_NH3lyase.
DR Pfam; PF00221; PAL; 1.
DR TIGRFAMs; TIGR01225; huth; 1.
DR PROSITE; PS00488; PAL_HISTIDASE; 1.
KW Lyase; Transferase.
SQ SEQUENCE 539 AA; 58138 MW; E0CE3B75B579B1D9 CRC64;
```

```
Query Match 30.5%; Score 814.5; DB 2; Length 539;
Best Local Similarity 40.3%; Pred. No. 4.5e-43;
Matches 202; Conservative 73; Mismatches 201; Indels 25; Gaps 8;
Qy 34 RIVLAPPARDCRASEAR--LGAVIREARHVYGLTGTGFGPLANRLISGENVRTLOANLVH 91
Db 32 RATVDVPAESAIAKAKSREIFEGIAEQNIPIYGVTTGYGEMIMQVDSKEVELQTNLVR 91
Qy 92 HLAGSGVPVLDWTTRAMVLAELVLSIAQASGASGTTARLIDLNSSELAPVPSRGTVG 151
Db 92 SHAGVGPLPAEDARAIVARLNTLAKGHSVRPILERLAQVLYNCGITPAIPEIGSLG 151
Qy 152 ASGDLTPLAHMVLCLQGRGDFDRDGTDLDAEGLRRGRKQPLDSLHSDALALVNGTSAM 211
Db 152 ASGD LAPLSHVA STLIGEG-VYLRDGRPVETAQVLAERGIEPLELRFKGLALINGTSGM 210
Qy 212 TGIALVNAHACHLGNWAVALTALLAECLRGTEAWAALS DL-RPHPGKQDAARLAR 270
Db 211 TGLGSLVVGRLAPQAQAEIVTALLIEAVRGSTSPFLAEGHDIAHPHEGQIDTAANMR 270
Qy 271 VDSARVVRHVAERRLDAGDIGTEPEAGQD-----AYSLRCAPQVLGAGPDTLA 320
Db 271 MRSGSLTVEHADRLRELQ-----KDEAGKDVORSEIYLOKAYSLRAIPQVVGAVR 325
Qy 321 WHDRVLTIELNAVTDNPVPPPDGSPALHGGNFMGQHVALTSDALATAVTTLAELAE 380
Db 326 HARHKLRIELNSANDNPLFPPEGKEI--FHGANFHGQPIAFAMDFVTTLTQLGVLAER 383
Qy 381 ARLTDERLNRLGPPFLHRGPAGLNSGFMGAQVATALLAEWRATGPAISHISTNAANQ 440
Db 384 NRYLNRLHSLGYPFLVSGDPGLHSGFAGAQYAPATAVAENRTTIGPASTOSVPSNGDN 443
Qy 441 VVSLGTIAARLCREKIDRWAEIILAILALCAQAEELRCGSG-LDGVSPACKLVOALRE 499
Db 444 VVSNGLISARNARRVLSNNKILAVEYLAQAQAVDI---SGRFDGLSPAAKAYEAVR 500
Qy 500 FPELETDRPLGOEIAALATHL 520
Db 501 VPTLGVDRYMADDIELVADAL 521
RESULT 3
Q4NR19_9DEL PRELIMINARY; PRT; 508 AA.
AC Q4NR19;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DE Histidine ammonia-lyase (EC 4.3.1.3).
GN ORFNames=AdhDRAFT_1286;
OS Anaeromyxobacter dehalogenans 2CP-C.
OC Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
OC Cyctobacterineae; Myxococaceae; Anaeromyxobacter.
OX NCBI_TaxID=290397;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=2CP-C;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RT "Sequencing of the draft genome assembly of Anaeromyxobacter
RT dehalogenans 2CP-C.";
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=2CP-C;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome assembly of Anaeromyxobacter
RT dehalogenans 2CP-C.";
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
```



```
RESULT 5
ID HUTH GLOVI STANDARD; PRT; 514 AA.
AC Q7NCB3;
DT 13-SEP-2005 (Rel. 48, Created)
DT 13-SEP-2005 (Rel. 48, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Histidine ammonia-lyase (EC 4.3.1.3) (Histidase).
GN Name=huth; OrderedLocusNames=glr3066;
OS Gloeobacter violaceus.
OC Bacteria; Cyanobacteria; Gloeobacteria; Gloeobacterales; Gloeobacter.
OX NCBI_TaxID=33072;
RN
RL NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=PCC 7421;
RX MEDLINE=22977040; PubMed=14621292;
RA Nakamura Y., Kaneko T., Sato S., Mimuro M., Miyashita H., Tsuchiya T.,
RA Sasamoto S., Watanabe A., Kawashima K., Kishida Y., Kiyokawa C.,
RA Kohara M., Matsumoto M., Matsumoto A., Nakazaki N., Shimo S.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of Gloeobacter violaceus PCC 7421, a
RT cyanobacterium that lacks thylakoids.";
RL DNA Res. 10:137-145(2003).
CC -1- CATALYTIC ACTIVITY: L-histidine = urocanate + NH(3).
CC -1- PATHWAY: Histidine degradation; first step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -1- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC which is formed autocatalytically by cyclization and dehydration
CC of residues Ala-Ser-Gly (by similarity).
CC -1- SIMILARITY: Belongs to the PAL/histidase family.
CC
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC
CC EMBL; BA00045; GAC91007.1; -; Genomic_DNA.
CC HSP; P21310; 1GKM.
CC HAMAP; MF 00229; -; 1.
CC InterPro; IPR005921; Huth.
CC InterPro; IPR001106; Phe/His_NH3lyase.
CC Pfam; PF00221; PAL; 1.
CC TIGRFAMs; TIGR01225; huth; 1.
CC DR PROSITE; PS00488; PAL_HISTIDASE; 1.
CC Complete proteome; Histidine metabolism; lyase.
FT MOD_RES 148 148 2,3-dihydroalanine (Ser) (By
FT similarity).
FT FT 147 149 5-imidazolinone (Ala-Gly) (By
FT similarity).
FT CROSSLNK 147 149
FT SEQUENCE 514 AA; 54037 MW; 15CC63A8ED444DA4 CRC64;
Query Match 29.9%; Score 797.5; DB 1; Length 514;
Best Local Similarity 38.9%; Pred. No. 5e-42;
Matches 198; Conservative 80; Mismatches 202; Indels 29; Gaps 10;
QY 22 LQAHAVASGARTVLAPARRDCRASEARIGAVIREARHVGLTTGFGPLANRLTSGEN 81
DB 17 VDLIVAVARGVPVRLSPASLELVRRSRFAVEALLEGDEIVYGLTTGFGFKNRRIPRSA 76
QY 82 VRTLQANLVHHLASGVGVPLDWTTRAMVLARLVSTIAQAGSAGSEGTIARLIDLNLSELA 141
DB 77 VEQLQQLNMSSAAGLCEPFGREVVRAMLLLRANTLAAQGYSGVRPETLQLLVAMLRGVH 136
QY 142 PAVPSRGTVCASGDLTFLAHVNLCLQGRGFLDRDTRLDGAGRLRGRLOPLDLHRDA 201
DB 137 PVVPCRGSVGSGDLAPLAHLALVLTGE- AEVGGEVLPFAAALARAGLEPIRLGAKG 195
QY 202 LALVNGTSGMTGALVNAHACRHGLGNVAVALTALLAECLRGRTGEMAAASDLRPHPGOK 261
DB 196 LALINGTQMSALGALTVHRAQRLAKLADACWLTLEATIGRSFAFLPHFRURPHPGQQ 255
QY 262 DAAARLRARVDGSAVRVVRHVAERRLDAGDIGTEPEAQDQDAYSLRCAPOVLGAGFDTLAW 321
```

```
Db 256 SSARNLLVLTEDSALIFASHAGCDR-----VQDAYSLRCAPOVHGASLDAISY 302
QY 322 HDRVLATIELNAVTDNP-VRPDDGVPALHGGNFMGQHVHVALTSDALATAVTVLVLGASLRQI 380
Db 303 AAGVIAIEINSVTDNPLIFADTQQV--VTGGHFGQPVAVASDVLAIALAELADISERRT 360
QY 381 ARLTDERLRNGLPFLHRRGPGAGLNSGFMGAQVTTATALLAEMRATG-PASIHSTSTNAANQ 439
Db 361 ERLVNADYSNGLPMLTEA--GGLHSGYVWAQYTAASLVSENKVLAAHPACVDSIPTSAQGE 419
QY 440 DVVSLGTIAARLCKREKIDRWAEILAILALCIAQAELRCGSLDGVSPAGKUL-VQALRE 498
Db 420 DHVSMGLTARKAVTVCDCNCEVIAELMCAQAQALDLR-GKLTPG---RGSRVGLEVIRA 475
QY 499 QFPLETRDPLGQFIAALAT-----HLIQ 522
Db 476 AVPHLESDDRVSRDIEKVVLMADGHLLE 504
RESULT 6
Q9KHJ9_9ACTO
ID Q9KHJ9_9ACTO PRELIMINARY; PRT; 523 AA.
AC Q9KHJ9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Putative phenylalanine ammonia lyase Encp.
GN Name=encp;
OS Streptomyces maritimus.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=115828;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Piel J., Hoang K., Moore B.S.;
RT "Natural Metabolic Diversity Encoded by the Enterocin Biosynthesis
RT Gene Cluster.";
RL J. Am. Chem. Soc. 122:5415-5416(2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=20579517; PubMed=11137817; DOI=10.1016/S1074-5521(00)00044-2;
RA Piel J., Hertweck C., Shipley P.R., Hunt D.M., Newman M.S.,
RA Moore B.S.;
RT "Cloning, sequencing and analysis of the enterocin biosynthesis gene
RT cluster from the marine isolate 'Streptomyces maritimus': evidence for
RT the derailment of an aromatic polyketide synthase.";
RL Chem. Biol. 7:943-955(2000).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=21942282; PubMed=11948863;
RX DOI=10.1002/1439-7633(20011001)2:10<784::AID-CBIC784>3.0.CO;2-K;
RA Hertweck C., Jarvis A.P., Xiang L., Moore B.S., Oldham N.J.;
RT "A mechanism of benzoic acid biosynthesis in plants and bacteria that
RT mirrors fatty acid beta-oxidation.";
RL Chembiochem 2:784-786(2001).
DR EMBL; AP254925; AAF81735.1; -; Genomic_DNA.
DR HSP; P21310; 1GKM.
DR GO; GO:0016211; F:ammonia ligase activity; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0009058; P:biosynthesis; IEA.
DR InterPro; IPR001106; Phe/His_NH3lyase.
DR Pfam; PF00221; PAL; 1.
KW Lyase.
SQ SEQUENCE 523 AA; 56416 MW; 012440F937D48FD9 CRC64;
Query Match 29.7%; Score 792.5; DB 2; Length 523;
Best Local Similarity 35.7%; Pred. No. 1.1e-41;
Matches 186; Conservative 101; Mismatches 217; Indels 17; Gaps 8;
QY 14 VELDRHIDLDOAHAVASGGARIVLAPARRDCRASEARIGAVIREARHVGLTTGFGPLA 73
DB 5 IELDMNVTLQLEBDAARQRTFVELSAPVRSVRASRDVLVKFVODERVIYGVNTSMGGFV 64
```

QY 74 NLSISENVRTLOANLVHLLASGVPLDWTTRAMVRLARLVSIQAAGSASEGTIARLI 133
 Db 65 DLHVPVQARQOENLINAATNGAYLDDTARTIMLSRIYSLARGNSAITPANLDDKLV 124
 QY 134 DLNSELAPVPSRGTVGASGDLTPLAHMVLCLQGRGDFLDRDGLDCAEGLRGRLOP 193
 Db 125 AVLNAGIIVPCIEKSGSLTSGDLGPLAALAVCAQWK-ARYNGQIMPGQALSEAGVEP 183
 QY 194 LLSHRDALVNGTSAMTGIALVNAHACRHGNVAVALTALLABCLRGRTTEMAAALSD 253
 Db 184 MELSFKDGLALNGTSGMVGLTGMVLQARRLVDRYLOVSALSVEGLAGMTKPFDPVRHG 243
 QY 254 LRPHPGQDAARLARARVDGSARVHVHIAERLDAGDIGTEPEAG-----QDAYSLRCAP 309
 Db 244 VAPHRGQVARSRLWEGLADSHLAVNELDTEQLT-AGEMGTAKAGSLAIEDAYSIRCTP 302
 QY 310 QVLGAGFTLAWHDRV---LTIELNAVTDNPFPPDGSVPALHGGNFMGQHVALTSDALA 366
 Db 303 QILGPVVDVL---DRIGATLQDELNSNDNPIVLPE-EAEVPHNGHFHQYVAMAMDHNL 358
 QY 367 TAVTVLAGLAERQIARLTDERLNRLGPPFLHRRGPAGLNSGFMGAQVTTALLAEMRA-TG 425
 Db 359 MALATVTNLANRRVDRFLDKNSNGLPAFLCREDPGLRLGLMGQGFMTASITAEITRTITI 418
 QY 426 PASIHSISNAANDVSLGTIAARLCREKIDRWAEIILAILALCLAAELRCGSGLDGV 485
 Db 419 PMSVQSLTSTADFQDIVSFGFVAARRAREVLNTAAAYVAVFELLCAQAVDIR---GADKL 475
 QY 486 SPAGKKLVQALREQPPPLETDRPLGOEIAALATHLLQSPV 526
 Db 476 SSFTRPLRYERTKIVPFFDRDETITDYVEKLAADLIAGEPV 516

RESULT 7

Q4H6E7_9DEIO PRELIMINARY; PRT; 499 AA.
 AC Q4H6E7_9DEIO
 DT 13-SEP-2005 (TRENBLrel. 31, Created)
 DT 13-SEP-2005 (TRENBLrel. 31, Last sequence update)
 DE Histidine ammonia-lyase (EC 4.3.1.3)
 GN ORFNames=DgeODRAFT_2894;
 OS Deinococcus geothermalms DSM 11300.
 OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
 OC Rhodocyclaceae; Deinococcus.
 OX NCBI_TaxID=319795;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=DSM 11300;
 RG US DOE Joint Genome Institute (JGI-PGF);
 RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,
 RA Hammon N., Israni S., Pitluck S., Richardson P.;
 RT "Sequencing of the draft genome assembly of Deinococcus geothermalms
 RT DSM 11300";
 RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=DSM 11300;
 RG US DOE Joint Genome Institute (JGI-ORNL);
 RA Laimer F., Land M.;
 RT "Annotation of the draft genome assembly of Deinococcus geothermalms
 RT DSM 11300";
 RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
 CC -! CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 CC EMBL; AAHE01000014; EAL61898.1; -; Genomic_DNA.
 KW Lyase.

Query Match 28.6%; Score 763; DB 2; Length 499;
 Best Local Similarity 40.0%; Pred. No. 7.2e-40;

Matches 200; Conservative 69; Mismatches 195; Indels 36; Gaps 12;
 QY 16 LDRHIDLQAHAVASGGGARIVLAPPARDRCRASEARLGAIVREARHVYGLTGTGFGFLANR 75
 Db 3 LDRQLTLDDFIRVRVGGEEVTLADAARTRMGARAVIERIVDGPBAVYGVNTGFGKFSV 62
 QY 76 LISGENVRTLOANLVHLLASGVPLDWTTRAMVRLARLVSIQAAGSASEGTIARLI 135
 Db 63 RVAREBELQOLHNLIVSHAIGVGAGLPAEVVRGMULLRAQSLALGHSGVVRPEVVELLLAL 122
 QY 136 LNSELAPVPSRGTVGASGDLTPLAHMVLCLQGRGDFLDRDGLDCAEGLRGRLOP 195
 Db 123 LNAGACPVVPAQSGVSGDLAPLAHLALALIGGGE-LEYGQVPRPADVLAEGLQPLIT 181
 QY 196 LSHRDALVNGTSAM---TGIALVNAHACRHGNVAVALTALLABCLRGRTTEMAAALSD 253
 Db 182 LEAKEGLALINGTQMGSLALLALHLDARTLLHTANLAAAMT---VEALSGSHRPFSEGVV 238
 QY 253 LRPHPGQDAARLARARVDGSARVHVHIAERLDAGDIGTEPEAGQDAYSLRCAPV 312
 Db 239 SLRPHPGALEVAADLRFLHGS-----IAPAHACGKV-----QDAYSLRAVPOVH 285
 QY 313 GAGFTLAWHDRVLTIELNAVTDNPFPPDGSVPALHGGNFMGQHVALTSDALATATVTV 371
 Db 286 GASLDALMQAGRVLEVEFASVTDNPLIIPETGEV---ISGPHGQPLAALADALKVAVAE 343
 QY 372 LAGLAERQIARLTDERLNRLGPPFLHRRGPAGLNSGFMGAQVTTALLAEMRA-TG-PASTH 430
 Db 344 LANISERSEQLNLPALS-GLPGFL-TPEGGLSSGFMIAQYTTAAALVSENKVLHAPASVD 401
 QY 431 SISITNAANDVSLGTIAARLCREKIDRWAEIILAILALCLAAELRCGSGLDGVSP 487
 Db 402 SIPTSANOEDHVSMAHGARGRLQILENAQSVISIELLCAAQALDFQSLRAGRQVQAA--- 459
 QY 488 AGKKLVQALREQPPPLETDR 507
 Db 460 -----YERIQEVAPLQDQR 474
 RESULT 8
 HUTH_AZOSE STANDARD; PRT; 526 AA.
 ID QSNZXB;
 AT 13-SEP-2005 (Rel. 48, Created)
 DT 13-SEP-2005 (Rel. 48, Last sequence update)
 DT 13-SEP-2005 (Rel. 48, Last annotation update)
 DE Histidine ammonia-lyase (EC 4.3.1.3) (Histidase).
 GN Name=huth; OrderedLocusNames=AZOSEA32610; ORFNames=ebsA5742;
 OS Azococcus sp. (strain EBN1).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Rhodocyclales;
 OC Rhodocyclaceae; Azococcus.
 OX NCBI_TaxID=761114;
 RN [1]
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
 RA PubMed=15551059; DOI=10.1007/s00203-004-0742-9;
 RA Rabus R., Kube M., Heider J., Beck A., Heitmann K., Widdel F.,
 RA Reinhardt R.;
 RT "The genome sequence of an anaerobic aromatic-degrading denitrifying
 RT bacterium, strain EBN1";
 RL Arch. Microbiol. 183:27-36 (2005).
 CC -! CATALYTIC ACTIVITY: L-histidine = urocanate + NH(3).
 CC -! PATHWAY: Histidine degradation; first step.
 CC -! SUBCELLULAR LOCATION: Cytoplasmic (Potential).
 CC -! PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
 CC which is formed autocatalytically by cyclization and dehydration
 CC of residues Ala-Ser-Gly (By similarity).
 CC -! SIMILARITY: Belongs to the PAL/histidase family.
 CC
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use as long as its content is in no way modified and this statement is not
 CC removed.

```
-----
CC EMBL; CR555306; CAI09386.1; -; Genomic_DNA.
DR HAMAP; MF_00229; -; 1.
DR InterPro; IPR005921; huth.
DR InterPro; IPR001106; Phe/His_NH3lyase.
DR Pfam; PF00221; PAL; 1.
DR TIGRFAMs; TIGR01225; huth; 1.
DR PROSITE; PS00488; PAL_HISTIDASE; 1.
KW Complete proteome; Histidine metabolism; Lyase.
FT MOD_RES 144 144
FT CROSSLINK 143 145
FT SEQUENCE 526 AA; 54655 MW; 94B1E777DC1574D3 CRC64;
Query Match 28.6%; Score 762.5; DB 1; Length 526;
Best Local Similarity 40.1%; Pred. No. 8.3e-40;
Matches 200; Conservative 65; Mismatches 211; Indels 23; Gaps 9;
Qy 20 IDLQAHAVASGGARIVLAPPARDRCRASEARLGAIVREARHVYLTGFGFPLANRLISG 79
Db 11 LTLAELRTIAPSDSKLELEPACFPVARGAATVAIARSGEPAVGINTGFGRLAQTHIPD 70
Qy 80 ENVETLQANLVHHLASGVGVLDWTTARAMVRLVLSIAQASGSGTTARLIDLINSE 139
Db 71 DQELLOKNLVLSHAGVGEPLSAPTIVRLVIALKIASLARGHSGVRMELINALIGLFNAG 130
Qy 140 LAPAVPSRGTVGASGDLTPLAHMVLCLQGRGDFLDRDGTLDGAEGLRRGRLOPLDLSHR 199
Db 131 VIPRVPKSGVSGASGDLAPLAHLASALLGIGEAY-VDGRHVPAATEALAIAGLAPMTLAAK 189
Qy 200 DALALVNGTSAMTGIALVNAHACHHLGNVAVALTALLAECLIRGRTAEAWAALSRLRPHG 259
Db 190 EGLALLNGTQVSTALAVNLFAITVFTALVAGALSVDAAAGSKFPDARIHALRGQPG 249
Qy 260 QXDAARLARVVGSGARVVRHVIAERLDAGDIGTEPEAGQDAYSRLCAQVILGAGDTL 319
Db 250 QIDAAATYRQLLEGSGINLAH-----RDCGKV-----QDPYSRLCQPVVMGACLDQM 296
Qy 320 AWHDRVLTIENAVTNP-VFPDGSVPALHGGFMGQHVALTSDALATATVTLAGLAER 378
Db 297 RHAARVLLIEANAVSDNPLFPDPSGEV--LSGGNFHGEPAFAADALALAAAEIGALAE 354
Qy 379 QIARLTDERLNRLGPPFLHRGPAGLNGFMGAQVATATALLAEMRATG-PASIHISISTNAA 437
Db 355 RIALLIIDATLS-GLPPFLVT-EGVNSGFMIAHVATAALASENKLIAHPASVDSLPISAN 412
Qy 438 NQDVVSLGTIAARLCREKIDRWAEILAILALCLAAQAAELRCGSLDGVSPAGKKLVQALR 497
Db 413 QEDHVSMTFAARKLGLADNTATILAILLAAAGVVELRAP---HRTSPRLQAVLALIR 469
Qy 498 EQPPLFTDPLRGQEIALL 516
Db 470 SRVPHYDIDRYFAPDIASI 488
RESULT 9
ID HUTH_STIAU STANDARD; PRT; 510 AA.
AC Q93TX3;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Histidine ammonia-lyase (EC 4.3.1.3) (Histidase).
GN Name=huth;
OS Stigmatella aurantiaca.
OC Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
OC Cystobacterineae; Cystobacteraceae; Stigmatella.
OX NCBI_TaxID=41;
RN [1]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RX MEDLINE=21110452; PubMed=11182319; DOI=10.1016/S1074-5521(00)00056-9;
RA Silakowski B., Nordaek G., Kunze B., Blocker H., Muller R.;
```

```
RT "Novel features in a combined polyketide synthase/non-ribosomal
RT peptide synthetase: the myxalamid biosynthetic gene cluster of the
RT myxobacterium Stigmatella aurantiaca Sgal5.,"
RL Chem. Biol. 8:59-69(2001).
CC -|- CATALYTIC ACTIVITY: L-histidine = urocanate + NH(3).
CC -|- PATHWAY: Histidine degradation; first step.
CC -|- SUBCELLULAR LOCATION: Cytoplasmic (potential).
CC -|- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC which is formed autocatalytically by cyclization and dehydration
CC of residues Ala-Ser-Gly (By similarity).
CC -|- SIMILARITY: Belongs to the PAL/histidase family.
CC
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC
CC EMBL; AF319998; AAK57183.1; -; Genomic_DNA.
DR HESP; P21310; IGKM.
DR HAMAP; MF_00229; -; 1.
DR InterPro; IPR005921; huth.
DR InterPro; IPR001106; Phe/His_NH3lyase.
DR Pfam; PF00221; PAL; 1.
DR TIGRFAMs; TIGR01225; huth; 1.
DR PROSITE; PS00488; PAL_HISTIDASE; 1.
KW Histidine metabolism; Lyase.
FT MOD_RES 146 146
FT CROSSLINK 145 147
FT SEQUENCE 510 AA; 54201 MW; DE78439DB4FBF3A6 CRC64;
Query Match 28.2%; Score 753; DB 1; Length 510;
Best Local Similarity 38.3%; Pred. No. 3.2e-39;
Matches 198; Conservative 85; Mismatches 202; Indels 32; Gaps 10;
Qy 7 MSPPKPAVELDRHIDLDQAHAVASGGARIVLAPPARDRCRASEARLGAIVREARHVYGLT 66
Db 1 MSRPRLNIDGDT-LKLEETILOVARHTVVELAPAAARVKAARDLVDRAAGDTPSGIN 59
Qy 67 TGFGLPLANRLISGENVRLTQANLVHHLASGVGVLDWTTARAMVRLVLSIAQASGASE 126
Db 60 TGFGLTAEVRIDKDLRELQNLILSHAAGVSGPLPLPEARVLLLLRCNVLAKGYSGIRP 119
Qy 127 GTTARLIDLINSELAPVPSRGTVGASGDLTPLAHMVLCLQGRGDFLDRDGTLDGAEGL 186
Db 120 ETLTALALEMLNRDVPVWPVPERGSGASGDLAPLAHLALVFIGEAFYK-GERLPAAQAL 178
Qy 187 RRGRLQPLDLSHRDALVNGTSAMTGIALVNAHACHHLGNVAVALTALLAECLIRGRTEA 246
Db 179 ERAGLKPVVLEAKEGLALVNGTQAMCAVGTILQLRAEMLADLADAGMTLEGLLSHKP 238
Qy 247 WAAALSRLRPHGQKQDAARLARVVGSGARVVRHVIAERLDAGDIGTEPEAGQDAYSRLR 306
Db 239 FIPEIQDVRAHEGQKACAAHRELLADLSALVESHVNCSE-----VQDPYSRLR 285
Qy 307 CAPQVLGAGPDTLAWHDRVLTIELNAVTDNP-VFPDGSVPALHGGFMGQHVALTSDAL 365
Db 286 CMPQVHGAAREGLSFARRILEVEINSATDNPVLFVETERI--VSGGNFHGQPVSLADVA 343
Qy 366 ATAVTVLAGLAERQIARLTDERLNRLGPPFLHRGPAGLNGFMGAQVATATALLAEMRA-T 424
Db 344 AMALTQLSAISERRVQLVNPALS-GLPPPLAKN-SGLNSGFMIAQVTSALVAESRVL 401
Qy 425 GPASIHISITNAANQDVVSLGTITAAARLCREKIDRWAEILAILALCLAAQAAE-----LRCGS 480
Db 402 HPASVDSIPSSAGREDHVSNGMTAALKRGQVADFTRSCLAIELLVAQAALDYQPTRAKG 461
Qy 481 GLDGVSPAGKKLVQALREQPPPLETORPLRQEIATAALA 517
Db 462 G-----PQAAAYELIRSKIPTMEKORELHRDIAAVS 491
```

```

RESULT 10
HUTH_BURMA STANDARD; PRT; 507 AA.
ID AC 62LJ6;
DT 13-SEP-2005 (Rel. 48, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Histidine ammonia-lyase (EC 4.3.1.3) (Histidase).
GN Name=huth; OrderedLocusNames=BMA0645;
OS Burkholderia mallei (Pseudomonas mallei).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia.
OX NCBI_TaxID=13373;
RN [1]

NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RP STRAIN=ATCC 23344;
RC PubMed=15377793; DOI=10.1073/pnas.0403306101;
RA Niernan W.C., Deshazer D., Kim H.S., Tettelin H., Nelson K.E.,
RA Feldblyum T.V., Ulrich R.L., Ronning C.M., Brinkac L.M.,
RA Daugherty S.C., Davidson I.D., DeBoy R.T., Dimitrov G., Dodson R.J.,
RA Durkin A.S., Gwinn M.L., Haft D.H., Khouri H.M., Kolonay J.F.,
RA Madupu R., Mohammed Y., Nelson W.C., Radune D., Romero C.M.,
RA Sarria S., Selengut J., Shamblin C., Sullivan S.A., White O., Yu Y.,
RA Zafar N., Zhou L., Fraser C.M.;
RT "Structural flexibility in the Burkholderia mallei genome.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:14246-14251(2004).
CC -!- CATALYTIC ACTIVITY: L-histidine = urocanate + NH(3).
CC -!- PATHWAY: Histidine degradation; first step.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -!- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC which is formed autocatalytically by cyclization and dehydration
CC of residues Ala-Ser-Gly (by similarity).
CC -!- SIMILARITY: Belongs to the PAL/histidase family.
CC
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC
CC EMBL; BMA0645; -; 1.
CC TIGR; CP000010; AAU49673.1; ALT_INIT; Genomic_DNA.
CC HAMAP; MF_00229; -; 1.
CC InterPro; IPR005921; Huth.
CC DR InterPro; IPR001106; Phe/His_NH3lyase.
CC PFam; PF00221; PAL; 1.
CC TIGRFAMs; TIGR01225; huth; 1.
CC DR PROSITE; PS00488; PAL_HISTIDASE; 1.
CC Complete proteome; Histidine metabolism; Lyase.
CC MOD_RES 142 142 2,3-didehydroalanine (Ser) (By
CC similarity).
CC CROSSLNC 141 143 5-imidazolinone (Ala-Gly) (By
CC similarity).
CC SEQUENCE 507 AA; 53141 MW; 77B4E49060B6D4AE CRC64;
Query Match 28.1%; Score 749.5; DB 1; Length 507;
Best Local Similarity 38.4%; Pred. No. 5.2e-39;
Matches 202; Conservative 74; Mismatches 213; Indels 37; Gaps 11;
QY 4 MLAMSPKPAVELDRHDLQAHAVAGSGARIVLAP---PARDRCRASEARLGNVREAR 60
DB 1 MITLTPGR-----LTLPLQRIARENVQIALDPASFAAIDRAQAQVADIAA---KGE 49
QY 61 HVYGLTGFGLANRLISGENVRTLOANLVHLLASGVGPVLDWTTARAVLARLVSTAQ 120
DB 50 PAYGINTGFGRLASTHPHOQLLELQKNLVLSHAGVGPSPARVPRVRLMALKLSSLGRG 109
QY 121 ASGASEGTIARLLDLNLSAPVPSRGTVGASGDLTFLAHMVLCLQGRGDFLDRDGTGL 180
DB 110 HSGIRRVVMDALVTLFNADVLPLIPVKGSGASGDLAPLAHMSAVLIGDVFIR-GERA 168
QY 181 DGAEGLRGRQLQDLDRDALALVNGTSAMTGIALVNAHACHRLGNVAVALTALLAECL 240

```

```

169 SAABGLRVAGLAPITLLEAKGGLALLNGTQASTALADNLFAIEDLYRTALVSGALSVDA 228
241 RGRTEAWAAALSDLRPHGPQKDAARLFRVVDGSRVVRHVIAERRLDAGDITGTEPEAQ 300
229 AGSVKPPDARITHELGRHGQIDAAAAYSLDGSAINVSH-----RCDKQV 275
301 DAYSIRCAPQVLGAGFTFLAWHDRVLTIELNAVTDNP-VFPDGSVPALHGNFMGQHV 359
276 DPLSLRCQPVMGACLDQIRHAAGVLLTEANAVSDNPLIFPDGTGV--LSGNGFHAEPVA 333
360 LTSALATATVTLAGLABRQIARLTDBELNGLPFLHGPAGLSGFMGQVQTATALLA 419
334 PAADNLATAAAEIGALAEIRIALLIDATLS-GLPPLVK-DGVNMSGFMIAHVTTAAALAS 391
420 EMRATG-PASIHSTISNAANOVDVSLGTIAARLCREKIDRWAEIILAILALCIAQAE 478
392 ENKTLAPASVDSLPTSANQEDHVSMTAFARKLTDIAENVANILAIELLAQAQVDLRA 451
479 GSGLDGVSPPAGKLVQALREQFPLETDPRPQBIALATHLLQOS 524
452 P---HATSPALQHAMKTIRADVAHYDLDPADPAPVAVRVRERA 494

RESULT 11
HUTH_BURPS STANDARD; PRT; 507 AA.
ID AC Q63SH6;
DT 13-SEP-2005 (Rel. 48, Created)
DT 13-SEP-2005 (Rel. 48, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Histidine ammonia-lyase (EC 4.3.1.3) (Histidase).
GN Name=huth; OrderedLocusNames=BPSL2344;
OS Burkholderia pseudomallei (Pseudomonas pseudomallei).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia; pseudomallei group.
OX NCBI_TaxID=28450;
RN [1]
NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=K96243;
RA Holden M.T.G., Titball R.W., Peacock S.J., Cerdano-Tarraga A.-M.,
RA Atkins T., Crossman L.C., Pitt T., Churcher C., Mungall K.L.,
RA Bentley S.D., Sebahia M., Thomson N.R., Bason N., Beacham I.R.,
RA Brooks K., Brown K.A., Brown N.F., Challis G.L., Cherevach I.,
RA Chillingworth T., Cronin A., Crossett B., Davis P., DeShazer D.,
RA Feltwell T., Fraser A., Hance Z., Hauser H., Holtroyd S., Jagels K.,
RA Keith K.E., Maddison M., Moule S., Price C., Quail M.A.,
RA Rabinowitsch E., Rutherford K., Sanders M., Simmonds M.,
RA Songsvilai S., Stevens K., Tumapa S., Vesaratchavest M.,
RA Whitehead S., Yeats C., Barrell B.G., Oyston P.C.F., Parkhill J.;
RT "Genomic plasticity of the causative agent of melioidosis,
RT Burkholderia pseudomallei.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:14240-14245(2004).
CC -!- CATALYTIC ACTIVITY: L-histidine = urocanate + NH(3).
CC -!- PATHWAY: Histidine degradation; first step.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -!- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC which is formed autocatalytically by cyclization and dehydration
CC of residues Ala-Ser-Gly (by similarity).
CC -!- SIMILARITY: Belongs to the PAL/histidase family.
CC
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC
CC EMBL; BX571965; CAH36347.1; -; Genomic_DNA.
CC HAMAP; MF_00229; -; 1.
CC InterPro; IPR005921; Huth.
CC DR InterPro; IPR001106; Phe/His_NH3lyase.
CC PFam; PF00221; PAL; 1.

```

```
DR  TIGRFAMS; TIGR01225; huth; 1.
DR  PROSITE; PS00488; PAL_HISTIDASE; 1.
KW  Complete proteome; Histidine metabolism; Lyase.
FT  MOD_RES 142 2,3-didehydroalanine (Ser) (By
FT  similarity)
FT  CROSSLINK 141 143 5-imidazolinone (Ala-Gly) (By
FT  similarity)
SQ  SEQUENCE 507 AA; 53111 MW; 77AFPE21CBAC651E CRC64;

Query Match      28.1%; Score 749.5; DB 1; Length 507;
Best Local Similarity 38.4%; Pred. No. 5.2e-39;
Matches 202; Conservative 74; Mismatches 213; Indels 37; Gaps 11;

Qy  4 MLAMSPKPAVELDRHLDLQAHAVAGGARIVLAP---PARDRCRASEARLGVAREAR 60
Db  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  1 MITLTPGR-----LTLPLQRIARENVQIALDPASFAAIDRGAQAVADIAA---KGE 49

Qy  61 HVYGLTTFGFGPLANRLISGENVRTLOANLVHHLASGVGVLDWTTRAMVRLVSTAQ 120
Db  ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db  50 PAYGINTGFGRLASTHPHQLELLQKNLVLSHAVGVGEPMARPVVRLMLKLSLGRG 109

Qy  121 ASGASECTIARLIDLLSELAPAVPSRGTVGASGLTPLAHMVLCLQGRGDFLDRDCTRL 180
Db  ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db  110 HSGIRRVMDALVTLFNADVPLIPVSGVSGASGLIAPLHMSAVLIGIGDVFIR-GERA 168

Qy  181 DGAEGLRRGLQPLDLSHRDALALVNGTSAVTGIALVNAHACRHLGNWVALTALLAEC 240
Db  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  169 SAAREGLRVAGLPLTLEAKGELLNGTQASTALALONLFAIEDLYRTALVSGALSVDAA 228

Qy  241 RGRTEAWAALSOLRPHPGKDAARLARVDSGARVVRVIAERRLDAGDIGTEPEAGQ 300
Db  ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db  229 AGSVKPPDARIHELGRHGQIDAAAAYRSLDGSAINVSH-----RDCDKVQ 275

Qy  301 DAYSLRCAPOVLGAGFTLAWHDRVLTIELNAVTDNP-VFPDGSVPALHCGNFMGOHVA 359
Db  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  276 DPLSLRCOPQVMGACLOIRHAAGVLLIEANVSDNPLIFPDTGEV--LSGGNPHAFSPA 333

Qy  360 LTPDALATAVTVLAGLAERQIARLTDLRNLGRGLPPFLHRGPAGLNSGFMGAQVTTATALLA 419
Db  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  334 FAADNLIAAAEIGALAEIRIALIDATLS-GLPPFLVK-DGWNVSGFMIAHTVTAALAS 391

Qy  420 EMRATG-PASIHSTVNAQNDVSVLGTIAARLCREKIDRWAEILAILALCLQAABELRC 478
Db  ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db  392 ENKTLAHPASVDSLPTSANQEDHVSMTAARLADIENAVANITALELAAAQGVDLRA 451

Qy  479 GSGLDGVSPPAGKLVQALRSQFPLETRPLGQFIAALATHLQQS 524
Db  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  452 P---HATSPALQAMKTIADVAHYDLDDHYFADPIAVVARRVRE 494
```

```
RESULT 12
HUTH TREDE
ID  HUTH TREDE STANDARD; PRT; 507 AA.
AC  Q73Q56;
DT  13-SEP-2005 (Rel. 48, Created)
DT  13-SEP-2005 (Rel. 48, Last sequence update)
DT  13-SEP-2005 (Rel. 48, Last annotation update)
DE  Histidine ammonia-lyase (EC 4.3.1.3) (Histidase).
GN  Name=huth; OrderedLocusNames=rDE0588;
OS  Treponema denticola.
OC  Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
OX  NCBI_TaxID=158;
RN  [1]
RP  NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC  STRAIN=ATCC 35405 / DSM 14222;
RX  PubMed=15064399; DOI=10.1073/pnas.0307639101;
RA  Seshadri R., Myers G.S.A., Tettelin H., Eisen J.A., Heidelberg J.F.,
RA  Dodson R.J., Davidsen T.M., DeBoy R.T., Fouts D.E., Haft D.H.,
RA  Selengut J., Ren Q., Brinkac L.M., Madupu R., Kolonay J.F.,
RA  Durkin S.A., Daugherty S.C., Shetty J., Shvartsbeyn A.,
RA  Gebregeorgis E., Geer K., Tsegaye G., Malek J.A., Ayodeji B.,
RA  Shatman S., McLeod M.P., Smajs D., Howell J.K., Pal S., Amin A.,
RA  Vashisth P., McNeill T.Z., Xiang Q., Sodergren E., Baca E.,
```

```
RA  Weinstock G.M., Norris S.J., Fraser C.M., Paulsen I.T.;
RT  "Comparison of the genome of the oral pathogen Treponema denticola
RL  with other spirochete genomes.";
RL  Proc. Natl. Acad. Sci. U.S.A. 101:5646-5651(2004).
CC  -I- CATALYTIC ACTIVITY: L-histidine = urocanate + NH(3).
CC  -I- PATHWAY: Histidine degradation; first step.
CC  -I- SUBCELLULAR LOCATION: Cytoplasmic (Periplasmic).
CC  -I- PFM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC  which is formed autocatalytically by cyclization and dehydration
CC  of residues Ala-Ser-Gly (By similarity).
CC  -I- SIMILARITY: Belongs to the PAL/histidase family.
CC  -----
CC  This Swiss-Prot entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use as long as its content is in no way modified and this statement is not
CC  removed.
CC  -----
DR  EMBL; AE017248; AAS11083.1; -; Genomic_DNA.
DR  HSSP; P21310; 1B8F.
DR  TIGR; TDR0588; -.
DR  HAMAP; MF 00229; -; 1.
DR  InterPro; IPR005921; Huth.
DR  InterPro; IPR001106; Phe/His_NH3lyase.
DR  Pfam; PF00221; PAL; 1.
DR  TIGRFAMS; TIGR01225; huth; 1.
DR  PROSITE; PS00488; PAL_HISTIDASE; FALSE NEG.
KW  Complete proteome; Histidine metabolism; Lyase.
FT  MOD_RES 146 146 2,3-didehydroalanine (Ser) (By
FT  similarity)
FT  CROSSLINK 145 147 5-imidazolinone (Ala-Gly) (By
FT  similarity)
SQ  SEQUENCE 507 AA; 54038 MW; 8269DC8F8F8463C CRC64;

Query Match      28.0%; Score 746; DB 1; Length 507;
Best Local Similarity 37.2%; Pred. No. 8.7e-39;
Matches 189; Conservative 85; Mismatches 204; Indels 30; Gaps 8;

Qy  11 KPAVELDRHLDLQAHAVAGGARIVLAPPARDCRASEARLGVAREARHVYGLTTFG 70
Db  ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db  4 KPTVTGSSLTIEDVAVARHGAELVLSADAKKIKDVKIVDDIVKSGFTVIGISGFG 63

Qy  71 PLANRLISGENVRTLOANLVHHLASGVGVLDWTTRAMVRLVSTAQASGASEGTIA 130
Db  ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db  64 ELSTVTITKQNGALQKNLILSHACGVGNFPREDIVRAIMLLRLNTHASGFGVTPSPVD 123

Qy  131 RLIDLINSELAPAVPSRGTVGASGLTPLAHMVLCLQGRGDFLDRDTRLDGAEGLRGR 190
Db  ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db  124 ILVDMNLKGVIPYVPEKSGLSGASGDLANLAHIALVMIGEKAY-YEGKLMEGKAALAKAG 182

Qy  191 LQPLDLSHRDALALVNGTSAVTGIALVNAHACRHLGNWVALTALLAECRLGRTEAWAAA 250
Db  ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db  183 LKPVVLSGKDKGLINGTIPVMSGIGALALHDAQLKKAANMGASLVFEAFEGITAAALDPR 242

Qy  251 LSDLRPHPGQKAAARLARVDSGARVVRVIAERRLDAGDIGTEPEAGQDAYSLRCAPO 310
Db  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  243 IHKSRPHKGQIDTAAFLKXMKGSS-----SINTRENDVQDPVTLRCVPQ 287

Qy  311 VLQAGFDTLAWHDRVLTIELNAVTDNP-VFPDGSVPALHCGNFMGOHVALTSDALATAV 369
Db  ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db  288 VH GASADAIAVVRKVLIEINAVTDNPLVFPDNDHV--ISGGNPHGQPIAITMDFLGI 345

Qy  370 TVLAGLAERQIARLTDLRNLGRGLPPFLHRGPAGLNSGFMGAQVTTATALLAEMRATG-PAS 428
Db  ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db  346 SELANISERIERLNVNPLQNGGLPAFLIEN-GVNSGFMIPQVTAASLVSENKVLAPAS 404

Qy  429 IHSISTVNAQNDVSVLGTIAARLCREKIDRWAEILAILALCLQAABELRCGSLDGVS 488
Db  ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db  405 VDSITSSGNKEDHVSMTAARLADIENAVANITALELAAAQGVDLRA-----GVK 458

Qy  489 GK---KLVMALRQFPPLPDLTDRPLQGEI 513
Db  ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db  459 GKGTGEMMKLIRKHISKVTEDRILYNDI 486
```



```
Db 172 DGLRVAGLAPUTLOAKGLALNGTQASTALDLMFAIEDLYRTALVAGALSVDAAAGS 231
Qy 244 TEAWAALSDLRPHGPQKDAARLARVDGSRVVRHVIARRRDLADGIGTEPEAGODAY 303
Db 232 VKPFDARIELHGRHQGQIDAASVRELLEGSF-----INQSHRDCKV-----QDPY 278
Qy 304 SLRCAPQVLGAGFTLAWHDVRLTIELNAVTDNP-VFPDPGVSVALRGHGMFGQHVALT 362
Db 279 SLRCOPQVMGACLDQMRHAADVLLVEANVSDNPLIFPDTGEV--LSSGNFHAEPVAF 336
Qy 363 DALATATVVLGAEARQIARLTDELNRGLPPFHLRGAGLNSFGMCAQVATATALLAEMR 422
Db 337 DNLALAAAEIGALAEIRIALIDATLS-GLPFPFLVK-DGGVNSGFMIAHVTAAALASENK 394
Qy 423 ATG-PASHSISTWAQNVDSVLGTIAARLCREKIDRWAEITLALCLQAELRCGSG 481
Db 395 TLAHPASVDSLFTSANQEDHVSMTAFKRLADTADNTKHTALELLAAAGGVDLRAPY- 453
Qy 482 LDGVSFAGKLVQALRQFPFPLETRPLGQIEAALA 517
Db 454 --HTSPKLPWMTETIRGKVAHYELDHVFPADIAVIA 487

RESULT 15
HUTH_BACSU
AC P10944; STANDARD; PRT; 508 AA.
DT 01-JUN-1989 (Rel. 11, Created)
DT 01-JUN-1989 (Rel. 11, Last sequence update)
DE 10-MAY-2005 (Rel. 47, Last annotation update)
DE Histidine ammonia-lyase (EC 4.3.1.3) (Histidase).
GN Name=huth; OrderedLocusNames=BSU39350;
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN NUCLEOTIDE SEQUENCE.
RP MEDLINE=88257040; PubMed=2454913;
RA Oda M., Sugishita A., Furukawa K.;
RT "Cloning and nucleotide sequences of histidase and regulatory genes in
RT the Bacillus subtilis hut operon and positive regulation of the
RT operon.";
RL J. Bacteriol. 170:3199-3205 (1988).
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=168 / BGSC1A1;
RX MEDLINE=95219089; PubMed=7704263;
RA Yoshida K.-I., Sano H., Seki S., Oda M., Fujimura M., Fujita Y.;
RT "Cloning and sequencing of a 29 kb region of the Bacillus subtilis
RT genome containing the hut and wapA loci.";
RL Microbiology 141:337-343 (1995).
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RP STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377; DOI=10.1038/36786;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Besieres P., Bolotin A., Borchert S.,
RA Borris R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.-K., Codani J.-J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Duesterhoeft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.-D., Ewington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.-Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guiseppe G., Guy B.-J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.-F., Itaya M.,
RA Jones L.-M., Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M.,
RA Klein C., Kobayashi Y., Koetter P., Koningsstein G., Krogh S.,
RA Kumano M., Kurita K., Lapidus A., Iardinois S., Lauber J.,
RA Lazarevic V., Lee S.-M., Levine A., Liu H., Masuda S., Mauel C.,
RA Medigue C., Meene N., Mellado R.P., Mizuno M., Moestl D., Nakai S.,
RA Noback M., Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B.,
```

```
RA Park S.-H., Parro V., Pohl T.M., Portetelle D., Porwollik S.,
RA Prescott A.M., Presecan E., Pujic P., Purnelle B., Rapoport G.,
RA Rey M., Reynolds S., Rieger M., Rivolta C., Rocha E., Roche B.,
RA Rose M., Sadaie Y., Sato T., Scanlan E., Schleich S., Schroeter R.,
RA Scoffone P., Sekiguchi J., Sekowska A., Seror S.J., Serror P.,
RA Shin B.-S., Soldo B., Sorokin A., Tacconi E., Takagi T., Takanashi H.,
RA Takemaru K., Takeuchi M., Tamakoshi A., Tanaka T., Terpetra P.,
RA Tognoni A., Tosato V., Uchiyama S., Vandenbol M., Vannier F.,
RA Vassarotti A., Viari A., Wambutt R., Wedler E., Wedler H.,
RA Weitzneger T., Winters P., Wipat A., Yamamoto H., Yamane K.,
RA Yasumoto K., Yata K., Yoshida K., Yoshikawa H.-F., Zumstein E.,
RA Yoshikawa H., Zanchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256 (1997).
CC -I- CATALYTIC ACTIVITY: L-histidine = urocanate + NH(3).
CC -I- PATHWAY: Histidine degradation; first step.
CC -I- SUBCELLULAR LOCATION: Cytoplasmic (potential).
CC -I- INDUCTION: By histidine.
CC -I- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC which is formed autocatalytically by cyclization and dehydration
CC of residues Ala-Ser-Gly (By similarity).
CC -I- SIMILARITY: Belongs to the PAL/histidase family.
CC
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC
CC EMBL; M20659; AAA22538.1; -; Genomic DNA.
CC EMBL; D31856; BAA06644.1; -; Genomic DNA.
CC EMBL; Z99124; CAB15971.1; -; Genomic DNA.
CC PIR; S18810; UFBESH.
CC HSP; P21310; IGKM.
CC Subtilisin; EG10667; huth.
CC HAMAP; MF 00229; -; 1.
CC InterPro; IPR005921; Huth.
CC InterPro; IPR001106; Phe/His_NH3lyase.
CC Pfam; PF00221; PAL; 1.
CC TIGRFAMS; TIGR01225; huth; 1.
CC PROSITE; PS00488; PAL_HISTIDASE; 1.
CC Complete proteome; Histidine metabolism; Lyase.
CC MOD_RES 142 142 2,3-didehydroalanine (Ser) (By
CC FT similarity).
CC FT CROSSLINK 141 143 5-imidazolinone (Ala-Gly) (By
CC FT similarity).
CC SQ SEQUENCE 508 AA; 55675 MW; 869C323BFCC318E0 CRC64;
Query Match 27.3%; Score 728.5; DB 1; Length 508;
Best Local Similarity 35.5%; Pred. No. 1.1e-37;
Matches 182; Conservative 94; Mismatches 208; Indels 29; Gaps 9;
Qy 20 IDLDQAHAVASGGARIVL-----APPARDRCRASEARLGAIVIRHARHYGLTGTGPGPL 72
Db 2 VTLDGSSLLTADVARVLDFEFAAASESMERVKSGRAAVERIVRDEKTIYGTGKGF 61
Qy 73 ANRLISEGNVRLTQANLVHHLASGCVPLDWTARTAMVLARLVSIAGQASGEGTIARL 132
Db 62 SDVLIQKEDSAALQLNLILSHACGVDPFPECVSRAMLLLRANALLKGFSGVRAELIEQL 121
Qy 133 IDLNLSELAPVPSRGTGVASGDLTPLAHMVLCLQGRGDFLDRDGLDGAELRGRLQ 192
Db 122 LAFNLKRVFPIVPOQSIGSGDLAPLSHALALIQGGEVF-FEGEMPMANTGLKKAGIQ 180
Qy 193 PLDLSHRDALALVNGTSAMTGIALVNAHACRHIGNWAVALTALLAECLRGRTFAMAAALS 252
Db 181 PVTLTSEGLALINGTOAMGVVYAEAKLAYQTERIASLTIEGLQGIIDAFDEDIH 240
Qy 253 DLRFPHGQKDAARLARVDGSRVVRHVIARRRDLADGIGTEPEAGODAYSIRCAPOVL 312
Db 241 LARGYEQIDVAERIRFYLSDSLTTSTQ--GELRV-----QDAYSLRCIPQVH 286
```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2005, 08:57:25 ; Search time 46 Seconds
(without alignments)
945.378 Million cell updates/sec

Title: US-10-621-826-3
Perfect score: 2668
Sequence: 1 VKMLAMSPKPAVELDRHI.....RPLQETALATHLLQSPV 526

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/1/iaa/5_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/6_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/H_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/PCUTS_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/RE_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	774.5	29.0	513	2	US-09-833-745-49
2	761	28.5	508	2	US-09-902-540-15192
3	749	28.1	513	2	US-09-833-745-52
4	749	28.1	513	2	US-09-833-745-58
5	728.5	27.3	508	2	US-09-833-745-40
6	724	27.1	513	2	US-09-833-745-46
7	721.5	27.0	513	2	US-09-833-745-44
8	717.5	26.9	513	2	US-09-833-745-50
9	710.5	26.6	513	2	US-09-833-745-56
10	706	26.5	504	2	US-09-833-745-65
11	706	26.5	516	2	US-09-833-745-41
12	695.5	26.1	678	2	US-09-949-016-10001
13	694.5	26.0	657	2	US-09-949-016-6174
14	690	25.9	513	2	US-09-833-745-45
15	687	25.7	513	2	US-09-833-745-64
16	687	25.7	525	2	US-09-540-236-2855
17	684.5	25.7	513	2	US-09-833-745-47
18	684.5	25.7	513	2	US-09-833-745-48
19	683	25.6	508	2	US-09-489-039A-13575
20	681	25.5	513	2	US-09-833-745-53
21	680.5	25.5	513	2	US-09-833-745-61
22	678.5	25.4	513	2	US-09-833-745-57
23	676.5	25.4	513	2	US-09-833-745-59
24	676.5	25.4	513	2	US-09-833-745-60
25	674	25.3	513	2	US-09-833-745-55
26	672.5	25.2	635	2	US-09-833-745-38
27	670	25.1	677	2	US-09-833-745-39

28	666.5	25.0	657	2	US-09-833-745-37	Sequence 37, Appl
29	662.5	24.8	657	2	US-09-833-745-36	Sequence 36, Appl
30	660.5	24.8	502	2	US-09-833-745-66	Sequence 66, Appl
31	660.5	24.8	513	2	US-09-833-745-5	Sequence 5, Appl
32	660.5	24.8	513	2	US-09-833-745-43	Sequence 43, Appl
33	654	24.5	515	2	US-09-328-352-5090	Sequence 5090, Ap
34	650	24.4	513	2	US-09-833-745-62	Sequence 62, Appl
35	647	24.3	513	2	US-09-833-745-51	Sequence 51, Appl
36	643	24.1	509	2	US-09-833-745-34	Sequence 34, Appl
37	631.5	23.7	511	2	US-09-833-745-10	Sequence 10, Appl
38	631.5	23.7	511	2	US-09-833-745-42	Sequence 42, Appl
39	626	23.5	511	2	US-09-833-745-35	Sequence 35, Appl
40	624	23.4	515	2	US-09-252-991A-33046	Sequence 33046, A
41	617	23.1	513	2	US-09-833-745-54	Sequence 54, Appl
42	616.5	23.1	520	2	US-09-252-991A-33049	Sequence 33049, A
43	584	21.9	513	2	US-09-833-745-63	Sequence 63, Appl
44	573	21.5	405	2	US-09-833-745-4	Sequence 4, Appl
45	559.5	21.0	404	2	US-09-833-745-9	Sequence 9, Appl

ALIGNMENTS

RESULT 1
US-09-833-745-49
; Sequence 49, Application US/09833745
; Patent No. 6939541
; GENERAL INFORMATION:
; APPLICANT: ROBERTS, JOSEPH
; APPLICANT: SETHURAMAN, NATARAJAN
; APPLICANT: MACALLISTER, THOMAS
; TITLE OF INVENTION: CLONING, OVEREXPRESSION AND THERAPEUTIC USE OF
; TITLE OF INVENTION: BIOACTIVE HISTIDINE AMMONIA LYASE
; FILE REFERENCE: 078728/0106
; CURRENT APPLICATION NUMBER: US/09/833,745
; CURRENT FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/197,770
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 49
; LENGTH: 513
; TYPE: PRT
; ORGANISM: Bacillus subtilis
; US-09-833-745-49

Query Match	29.0%	Score 774.5;	DB 2;	Length 513;
Best Local Similarity	36.4%	Pred. No. 3.5e-67;		
Matches 187;	Conservative	94;	Mismatches 206;	Indels 27; Gaps 7;
Qy	20	IDLQAHAVASGGARIYL-----APPARDRCRASEARLGAIVREARHVVYLTGTGQPL	72	
Db	7	VTLGGSLTTADVARVLDFEEAAASESMERVKSRAAVERIVRDEKTIYINTGFGKF	66	
Qy	73	ANRLISEGNTVTLQANLVHHLASGVPLDWTARAVMLARLVLSIAQASGASEGTIARL	132	
Db	67	SDVLQKEDSAALQLNLILSHACGVGDPFPFECVSRAMLLLRANALLKFGSVRAELIEQL	126	
Qy	133	IDLLNSLAPVPSRGTVGASGDLTPLAHMVLCLQGRGDFLDRDGTLDGAEGLRRGLQ	192	
Db	127	LAFINKRVHVPVFOQSGISGSDLPUSHLALALIGQEVDFEGERMPANTGLKKAGIQ	186	
Qy	193	PLDLSHRDALALVNGTSAMTGIALVNAHACHRLGNNAVALTALAECLRGRTAEWAALS	252	
Db	187	PVTLTSKEGLALINGTOAMTAMGVVYAEAKLAYQTERIASLTIEGLQGIIDAFDEDIH	246	
Qy	253	-DRPHPGQKDAARLARVRDVGSGARVVRVIAERLRDAGDIGTEPEAGQDAYSURCAPOV	311	
Db	247	LALRGYQEQIDVAERIRFYLSDSGLTTSRQGLRV-----QDAYSURCIPQV	294	
Qy	312	LGAGFDTLAHDVRLVLTIELNAVTDNVPFPDGSVPALHGGNFMGSHVALTSDALATAVTV	371	
Db	295	HGATWTQLGYVKLEIEMNAATDNPLIFNDGDV--ISSGNFHQPIAFAMDFLKIAISE	352	


```
; APPLICANT: ROBERTS, JOSEPH
; APPLICANT: SETHURAMAN, NATARAJAN
; APPLICANT: MACALLISTER, THOMAS
; TITLE OF INVENTION: CLONING, OVEREXPRESSION AND THERAPEUTIC USE OF
; TITLE OF INVENTION: BIOACTIVE HISTIDINE AMMONIA LYASE
; FILE REFERENCE: 078728/0106
; CURRENT APPLICATION NUMBER: US/09/833,745
; CURRENT FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/197,770
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 58
; LENGTH: 513
; TYPE: PRT
; ORGANISM: Thermoplasma acidophilum
US-09-833-745-58

Query Match          28.1%; Score 749; DB 2; Length 513;
Best Local Similarity 37.4%; Pred No. 1.le-64;
Matches 199; Conservative 91; Mismatches 194; Indels 48; Gaps 15

QY      12   PAVELD-RHIDLDAQHAVASGGARIVLAPPARDRCRASEARLGA---VIREARHYVGLTT 67
        :|::||:::||::||::||::||::||::||::||::||::||::||::||::||::||:
Db       5   PMIEIDGRSLRVEDVYAVAYEYDRVSISD--DTLKAVEEKHEAFLKLNSKTYGVNT 61

QY      68   GGCPLEANRLLSGENVRTLOANLVHLHLAGCVGPLDWTTRAWVLARLVSIQAQSGASEG 127
        :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db      62   GFSGLLNVNIHQEIQLKNLRISHSSGGDYLENRYVRIMAVRLNLSLAAGYSAVSAD 121

QY     128   TIARLIDLLNSELPAPVPSRGTCVASGDLTPLAHMVLCLQGRGDFLDRDGTFLDCAEGLR 187
        :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db    122   LNNVMVEMLNRDVIPVPKYGSVGSGDLAPLANHIGLAMMGEGRAFDPEGRLMDSARALE 181

QY     188   RGRILQPIDLSDRALALVNQTSAMTG---IALVNAHACRHIGNNAVALTALLAECLRGT 244
        :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db    182   KAGLKPYQPKKEGVALINGTSPFMSGILSIAMDAHDI--LEN-AIRSALLSFELAGGTS 238

QY     245   EAWAA--ALSDLRPHPGOKDAARLARVDGSGARVVHVIAERRLDDAGDIGTEPEGAGDAY 303
        :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db    239   KAFTPWILGALRPILHGQVAIGNRFYITGTS-----DIVASKRADSVKV-----QDAY 286

QY     304   SLRCAPQVLGAGFDFTLAWHDRLVTIELNAVTDNF-VFPDPDGSPALHGNGFMGHVALTS 362
        :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db    287   TLRCIPQVGSVADVITYVENVLVSVEINSATDNPLVF--NGEEVVSNGFNHFGEPAVAAA 343

QY     363   DALATAVTIVLAGLAERQIARLTDERLNGLPPFLHRGPAGLNSFGMGAOVQTATALLAEMR 422
        :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db    344   DFLAIALTDLGNMVVERRIARLVDTNLSRGLPFFL-TPDSGLNSGYMTIPQYTAAALCNRNK 402

QY     423   ATG-PASIHGISITWAAQDVVSLGTIAARLCREKIDRWAEITAILALCAQAAELRCGSG 481
        :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db    403   VLAYPSSADIPTPSANQEDHSVGMGATCSKLLEIIDNVRYIIAYELLGSQLAF--TD 459

QY     482   LDGVSPAGKKLVQALRSQFPPLETDRPLG-----QEIAALATHL 520

Db    460   KLWGSPSTRKIYEKIRKVKELDHDRPPSFDIETIRKMDCKKETISALPAHL 511
```

RESULT 5

```

RESOL 3
US-09-833-745-40
; Sequence 40, Application US/09833745
; Patent No. 6939541
; GENERAL INFORMATION:
; APPLICANT: ROBERTS, JOSEPH
; APPLICANT: SETHURAMAN, NATARAJAN
; APPLICANT: MACALLISTER, THOMAS
; TITLE OF INVENTION: CLONING, OVEREXPRESSION
; TITLE OF INVENTION: BIOACTIVE HISTIDINE
; FILE REFERENCE: 078728/0106
; CURRENT APPLICATION NUMBER: US/09/833,745
; CURRENT FILING DATE: 2001-04-13
; PRIORITY APPLICATION NUMBER: 60/197,770

```

```

; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 40
; LENGTH: 508
; TYPE: PRT
; ORGANISM: Bacillus subtilis
; US-09-833-745-40

Query Match          27.3%; Score 728.5; DB 2; Length 508;
Best Local Similarity 35.5%; Pred. No. 1.le-62;
Matches 182; Conservative 94; Mismatches 208; Indels 29; Gaps 9

Qy 20 IDLDQAHAVASGGARIVL-----APPARDRCRASEARLGAIVREARHVYGLTTGFGPGL 72
Db 2 VTLDGSSLTADVARVLFDPEAAASAESEMERVKRAAVERIVRDEKTIYIGINTFGCGKF 61
Qy 73 ANRLISGENVRTTLQANLVHHLASGVPLDWTTCARAVMLRLVLSAQCASGASEGTIARL 132
Db 62 SDVLIQKEDSAALQNLIHLSHACGVGDPFPCEVSRAMLLLRANALLKFGSGVRAELIQ 121
Qy 133 IDLNSELAPAVPSRGTVGASGDLTPLAHVMVLCIQRGDFLDNRDGTRLDGAELRRGRILQ 192
Db 122 LAFLNKRVPVPIQOQSLGASGDLAPLSHLALALIGQGEVF-FEGERPAMTGLKKAGIQ 180
Qy 193 PLDSLHRDALVNGTSMGTGIALVNAHACHRLGNWAVALTALLAECLRGRTTEAWAAALS 252
Db 181 PVLTSKEGLALINGTQAMTGAVVAYTEAKLAYQTERIASLTIEGLQGIIDAPEDEIH 240
Qy 253 DLPHPCQKDAARLRARVDGSARVRHVIAERRLDAGDIGTEPEAGQDAYSRLCAPOVL 312
Db 241 LARGYQEQIDVAERIRFVLSDSGLTTSQ--GELRV-----QDAYSLRCIPQVH 286
Qy 313 GAGPDTTLAWHDRVLTIELNAVTDNPVPPDGSVPALHGGNFMGQHVALTSDALATAVTVL 372
Db 287 GATWOTLLGYVKEKLEIEMNAATDNPLIFNDGD-KVISGNGFHGQPIAFAMDFLKIAISEL 345
Qy 373 AGLAERQIARLTBERLNRGLPPFLHRGPAGLNSFGMAQVTTATALLAEMRATG-PASTHS 431
Db 346 ANTAERIERLVNPQLN-DLPPFLSPHP-GLQSGAMINQYAAASLVSENKTLAHFASVDS 403
Qy 432 ISTNAAMQDVVSLGTIARLCREKIDRWAEITLAILALCLAAELRCGSLDGVSPAGKK 491
Db 404 IPSANQEDHVSMTGIARHAYQVIANTRRVIAETALCALQAVEYR---GIEHAASYTKQ 460
Qy 492 LVQALREQFPFPLETRPLGQGEIAALATHLLQQS 524
Db 461 LFOEMRKVPVSIODRVFSYDIBRLTDLWKES 493

```

RESULT 6

US-09-833-745-46
; Sequence 46, Application US/09833745
; Patent No. 6939541
; GENERAL INFORMATION:
; APPLICANT: ROBERTS, JOSEPH
; APPLICANT: SETHURAMAN, NATARAJAN
; APPLICANT: MACALLISTER, THOMAS
; TITLE OF INVENTION: CLONING, OVEREXPRESSION AND THERAPEUTIC USE OF
; TITLE OF INVENTION: BIOACTIVE HISTIDINE AMMONIA LYASE
; FILE REFERENCE: 078728/0106
; CURRENT APPLICATION NUMBER: US/09/833,745
; CURRENT FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/197,770
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 46
; LENGTH: 513
; TYPE: PRT
; ORGANISM: Deinococcus radiodurans
US-09-833-745-46


```
Db 194 EGIALLNGTOASTAFALBGLFVAEDLFASATVCGAMSV-----AALGSRPPDP 243
Qy 257 -----HPGKDAARARVDSARVVRHIAERLDRAGDGTPEAGODAYSIRCA 308
Db 244 RIHRVVRGHTQMDAAT-----AYRHLVSSEIGOSHNSCEGKV-QDPYSLRCQ 291
Qy 309 PQVLGAGDTLAWDRVLTIELNAVTPVPDPDGSVPALHGGNFMGQHVALTSDALATA 368
Db 292 PQVWGACIQOTRSAAEVLEVEANSVDNPLVPADGDI--ISGNNFHAEPVVAADNLTALA 349
Qy 369 VTVLAGLAERQIARTLDERLNRGLPPFLHRGPAGLNSGFMGAQVTTATALLAEMRATG-PA 427
Db 350 IAEIGSLSERMALLIDSALSCKLPPFLVDN-GGVNSGFMIAQVTTAAALASENKTALHPA 408
Qy 428 SIHSISTWAAQDVVSLGTIAARLREKIDRWAEILAILALCLAAQAEIRCGSLDGVSP 487
Db 409 SVDSLTSANQEDHVSMTAARLRDMGENTRGILAVEYLAAAGQLDFRAPL---KSSP 465
Qy 488 AGKKLVQALRQFPLETRPLGQEI-----AALATH 519
Db 466 RIEARQILREKVPYDKRYFADPIEKANALLAQLAVH 504

RESULT 9
US-09-833-745-56
; Sequence 56, Application US/09833745
; Patent No. 6939541
; GENERAL INFORMATION:
; APPLICANT: ROBERTS, JOSEPH
; APPLICANT: SETHURAMAN, NATARAJAN
; APPLICANT: MACALLISTER, THOMAS
; TITLE OF INVENTION: CLONING, OVEREXPRESSION AND THERAPEUTIC USE OF
; FILE REFERENCE: 078728/0106
; CURRENT APPLICATION NUMBER: US/09/833,745
; CURRENT FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/197,770
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 56
; LENGTH: 513
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-833-745-56

Query Match 26.6%; Score 710.5; DB 2; Length 513;
Best Local Similarity 36.8%; Pred. No. 6.8e-61;
Matches 190; Conservative 82; Mismatches 222; Indels 23; Gaps 9;

Qy 10 PKPAVELDRHLDLQAHAVASGGARIVLAPPARDRCRASEARLGAIVREARHVYGLTTGF 69
Db 5 PEKVIELDGLTDELVN-LGKGRYKIKLTPTAEKRQVKSREVIDSIIEKTVWYGITTF 63
Qy 70 GPLANRLISGENVRTLOANTVHLASGVPVLDWTTTARAMVLARLVSTIAQASGASGEGTI 129
Db 64 GKPAATRTVIPKQLQVNLVRSSSHSGVKGKPLSPERCMLLARINVLAKYSGISLETL 123
Qy 130 ARLIDLNSLAPVPSRGTVGASGDLTPLAHMWLCLQGRGDFLDRDGTLDGAEGLRRG 189
Db 124 KQVLEMFNASCPLVPYKGTGASGDLAPLSHLALGLVGEKWKSPSGWADAKVYLEAH 183
Qy 190 RLQPLDLSHRDALALVNGTSAMTGIALVNAHACHRLGNWVALTALLAECLRLGRTEAWAA 249
Db 184 GLKPVILKPEGLALINGTOMITSLGCEAVERASAIARQADIVAALTLEVLKGTTKAFDT 243
Qy 250 AL-SDLRPHQOKDAARLARVDSARVVRHIAERLDRAGDGTPEAGODAYSIRCA 308
Db 244 DIHAPLRPHRQOI EVAFRFRSLLSDSEIAESHRFCDSRV-----QDAYTLRCC 291
Qy 309 PQVLGAGDTLAWDRVLTIELNAVTPVPDPDGSVPALHGGNFMGQHVALTSDALATA 368
Db 292 PQVWGACIQOTRSAAEVLEVEANSVDNPLVPADGDI--ISGNNFHAEPVVAADNLTALA 349
```

```
Qy 369 VTVLAGLAERQIARTLDERLNRGLPPFLHRGPAGLNSGFMGAQVTTATALLAEMRA-TGPA 427
Db 350 ITHLAAISERRIERLCNPSLSRELPAFL-VAEGGLNSGFMIAHCTAAALVSENKALCHPS 408
Qy 428 SIHSISTWAAQDVVSLGTIAARLREKIDRWAEILAILALCLAAQAAE-LRCGSLDGV 486
Db 409 SVDSLTSANQEDHVSMTAARLREKIDRWAEILAILALCLAAQAAE-LRCGSLDGV 486
Qy 487 PAGKKLVQALRQFPLETRPLGQEI AALATHLLQQ 523
Db 466 PL-EKVVDLVRVVRPWIKDRFMAPDIEAAHRLLEQ 501

RESULT 10
US-09-833-745-65
; Sequence 65, Application US/09833745
; Patent No. 6939541
; GENERAL INFORMATION:
; APPLICANT: ROBERTS, JOSEPH
; APPLICANT: SETHURAMAN, NATARAJAN
; APPLICANT: MACALLISTER, THOMAS
; TITLE OF INVENTION: CLONING, OVEREXPRESSION AND THERAPEUTIC USE OF
; FILE REFERENCE: 078728/0106
; CURRENT APPLICATION NUMBER: US/09/833,745
; CURRENT FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/197,770
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 65
; LENGTH: 504
; TYPE: PRT
; ORGANISM: Streptomyces griseus
US-09-833-745-65

Query Match 26.5%; Score 706; DB 2; Length 504;
Best Local Similarity 38.6%; Pred. No. 1.8e-60;
Matches 192; Conservative 69; Mismatches 212; Indels 24; Gaps 8;

Qy 27 AVASGGARIVLAPPARDRCRASEARLGAIVREARHVYGLTTGFGLPLANRLISGENVRTLQ 86
Db 15 AVARHGARVELSAAVAALAAARLIVDALAAKPFVGVSTGFGALASRHLGTTELRAQLQ 74
Qy 87 ANLVHHLASGVPVLDWTTTARAMVLARLVSTIAQASGASGASGEGTIARLIDLNSLAPVPS 146
Db 75 RNVIRSHAAGMGPRVEREVVRALMFLRLKTVASGHTGVRPEVAQTMDVNLNAGITPVVHE 134
Qy 147 RGTVGASGDLTPLAHMWLCLQGRGDFLDRDGTLDGAEGLRRGLQPLDLSHRDALALVN 206
Db 135 YGSLGSGDLAPLSHCLATLMGEAGEAGPDGTVRPAGELLAHAAGIAPVELREKEGLALLN 194
Qy 207 GTSAMTGIALVNAHACHRLGNWVALTALLAECLRLGRTEAWAAALSDLRPHQOKDAAR 266
Db 195 GTDGMGLMWALADLRNLYTSDITAALESALGLTQVLAPELHAIRPHPGQVSDN 254
Qy 267 LRARVDSARVVRHIAERLDRAGDGTPEAGODAYSIRCAPOVLGAGFTFLAWHDRV 326
Db 255 MSRVLAGSGLTGHH-----QDDAPRV-----QDAYSVRCAPQVNGAGRDTLDHAALVA 302
Qy 327 TIELNAVTDNVPFPDGSVPALHGGNFMGQHVALTSDALATVTVTLAGLAERQIARTLDE 386
Db 303 GRELASSVDNVPVLPDGRVES--NGNFHGFAPVAVYLDLFAITVAADLGSICERRTRDLRDK 360
Qy 387 RLNRGLPFLHRGPAGLNSGFMGAQVTTATALLAEM-RATGPASHTSITNAANDVVSIG 445
Db 361 NRSHGLPFL-ADDAGVDSGLMIAQYTOAALVSEKRLAVPASADSPSSAQWEDHVSIG 419
Qy 446 TIAARLREKIDRWAEILAILALCLAAQAEIRCGSLDGVSPAGKKLVQALR-----BOFP 501
Db 420 WSAARKLRTAVDNLARIVAVELYAATRAIELRAAEGLT-PAPASEAVVAALRAAGAE 478
```

```
Qy 502 PLETRPLGQEIAT 518
Db 479 P---DRELPDLAAADT 492

RESULT 11
US-09-833-745-41
; Sequence 41, Application US/09833745
; Patent No. 6939541
; GENERAL INFORMATION:
; APPLICANT: ROBERTS, JOSEPH
; APPLICANT: SETHURAMAN, NATARAJAN
; APPLICANT: MACALISTER, THOMAS
; TITLE OF INVENTION: CLONING, OVEREXPRESSION AND THERAPEUTIC USE OF
; TITLE OF INVENTION: BIOACTIVE HISTIDINE AMMONIA LYASE
; FILE REFERENCE: 078728/0106
; CURRENT APPLICATION NUMBER: US/09/833,745
; CURRENT FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/197,770
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 41
; LENGTH: 516
; TYPE: PRT
; ORGANISM: Streptomyces griseus
US-09-833-745-41

Query Match 26.5%; Score 706; DB 2; Length 516;
Best Local Similarity 38.6%; Pred. No. 1.9e-60;
Matches 192; Conservative 69; Mismatches 212; Indels 24; Gaps 8;

Qy 27 AVASGGARIVLAPPDRCRASEARLCAVIREARHVYGLTTGFGPLANRLISGNVRLT 86
Db 20 AVAHGARVELUSAAVEAALAAARLIVDALAKPEPVYGVSTGFGALASRHIGTELRAQLQ 79

Qy 87 ANLVHHLASGVGPVLDVTTARAVMLARLVSAQAQASGASEGTTARLDLNSLAPAVPS 146
Db 80 RNIVRSHAAAGVPRVEREVVRLMFLRKTVASGHTGVRPEVAQTMADVLNAGITPVVHE 139

Qy 147 RGTVGASGDLTPLAHMWLCTQGRGDFLDRCGLDGLAEGLRRGLQPLDLSHRDALVN 206
Db 140 YGSLGCSGDLAPLSHCALTLMGEAGEBPGDQVTRPAGELLAHAAGIAPVELREKEGLALLN 199

Qy 207 GTSAMTGIALVNAHACHRLGNWALVALLAECLRGRTAEWAALSLDLRPHPGOKDAAR 266
Db 200 GTDGLMGLVWALADLRNLVTSADITAALESLEALGTDKVLAPELHAIIRPHPGQVSADN 259

Qy 267 LRARVDSARVVRHVIAERRLDAGDIGTEPEAGQDAYSRLCAPQVLGAGFDTLAWHDRV 326
Db 260 MSRVLAGSGLTGH- - - - -QDDAPRV- - - - -QDAYSVRCAPQVNGAGRDTLDHAALVA 307

Qy 327 TIENAVTDNPPVPPDGSVPALHGNFMGQHVLTSDALATAVTVLAGLAEARQIARLTDE 386
Db 308 GRELASSVDNPPVLPDGRVES- - -NGNFHGPAPVAVVLDFLAIVAADLGSICERRTRDLLDK 365

Qy 387 RLNRGLPPELHRCGAGNSGPMGAQVATATLAEM- - -RATGPASTHSTNAANOQDVVSLG 445
Db 366 NRSGLPPEFL- - -ADGAVDSGLMIAQYTOAALVSEMKRLAVPASADSIPSSAMQEDHVS 424

Qy 446 TIAARLCREKIDRWAEILALCLAAQAEIRCGSLDGVSPAGKLVQALR- - - - -EQFP 501
Db 425 WSAARKLRTAVDNLARIVAVELYAATRAIELRAAEGLT- - -PAPASEAVVAALRAAGSGPG 483

Qy 502 PLETRPLGQEIAT 518
Db 484 P---DRELPDLAAADT 497

RESULT 12
US-09-949-016-10001
; Sequence 10001, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14

Qy 502 PLETRPLGQEIAT 518
Db 479 P---DRELPDLAAADT 492

RESULT 13
US-09-949-016-6174
; Sequence 6174, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14

; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14

Query Match 26.1%; Score 695.5; DB 2; Length 678;
Best Local Similarity 35.7%; Pred. No. 3.2e-59;
Matches 193; Conservative 81; Mismatches 227; Indels 39; Gaps 11;

Qy 6 AMSP- - - - -PKPAVELD- - -RHIDLDQAHAVASGGARIVLAPPDRCR 46
Db 109 AMSPDFIPSPQEGVLYSKYREPEKYEILDGDRLTEDLVNLGKGRYKIKLTPTAEKRVQ 168

Qy 47 ASBARLAVIREARHVYGLTTGFGPLANRLISGNVRLTQANLVHHLASGVGPVLDVTTA 106
Db 169 KSREVIDSIIEKTVVYGVITGFGKFAITVPIKQLQELQVNLVRSHSSGVKPLSPERC 228

Qy 107 RAMVLARLVSAQAQASGASEGTTARLDLNSLAPAVPSRGTVGASGDTPLAHMWLCL 166
Db 229 RMLLAURINVLAKGYSGISLETILKQVIEMFNASCLPVYPEKTVGASGDLAPUSHLALGL 288

Qy 167 QCRGDFLDRCGLDGLAEGLRRGLQPLDLSHRDALALVNGTSAMTGIALVNAHACHRLG 226
Db 289 VGEKMWSPKSGWADAKYVLEAHGLKPVILKPEGLALINGTQMITSLGCEAVERASAIA 348

Qy 227 NWAVALTALLAECLRGRTAEWAALSLDLRPHPGOKDAARLRAKRVDSARVVRHVIAERR 286
Db 349 ROADIVAALTLEVLKGTTKAFDTDIHALRPHRGQIEVAFRFRSLDSDHHPSEIAESHRF 408

Qy 287 LDAGDIGTEPEAGQDAYSRLCAPQVLGAGFDTLAWHDRVLTIELNAVTDNP- - -VFP 345
Db 409 CD- - - - -RVQDAYTLRCCPQVHGVVNDTIAFVKNIITTELSATDPMVFNARGE- 458

Qy 346 PALHGNFMGQHVLTSDALATAVTVLAGLAEARQIARLTDERLNRGLPPELHRCGPAGLNS 405
Db 459 - - -TISGNGFHEGYPAKALDYLAIGIHELAAISERRIELCNPSLSE- - -LPAFL- - -VAEGGLNS 515

Qy 406 GFPGAQVATALLAEWRA- - -TGPASIHISISTNAANOQDVVSLGTTAARLCREKIDRWAEILA 464
Db 516 GFPIAHCTAAALVSENKALCHPSVSDLSLSTSAATEDHVSVMGWAARKALRVIEHVEQVLA 575

Qy 465 ILALCLAAQAE- - -LRCSGLDGVSPAGKLVQALREQPPELTDRLPGLQETAAALATHLLQ 523
Db 576 IELLAACQGIETFLR- - - - -PLKTTTFL- - -EKVYDLVRSVVRPWIKORFMADPIEAHRLLEQ 631

; Sequence 6174, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2005, 09:05:35 ; Search time 165 Seconds
(without alignments)
1331.989 Million cell updates/sec

Title: US-10-621-826-3

Perfect score: 2668

Sequence: 1 VKPMLAMSPKPAVELDRHI.....RPLQGEIATLATHLLQSPV 526

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA Main:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2668	100.0	526	US-10-621-826-3	Sequence 3, Appli
2	2665	99.9	526	US-11-069-633-8	Sequence 8, Appli
3	2444	91.6	482	US-10-369-493-7816	Sequence 7816, Ap
4	1335	50.0	540	US-10-464-609-4	Sequence 4, Appli
5	837	31.4	529	US-10-369-493-8812	Sequence 8812, Ap
6	774.5	29.0	513	US-09-833-745-49	Sequence 49, Appl
7	755	28.3	501	US-10-369-493-4756	Sequence 4756, Ap
8	755	28.3	501	US-10-369-493-7515	Sequence 7515, Ap
9	755	28.3	507	US-10-282-122A-49837	Sequence 49837, A
10	753.5	28.2	501	US-10-369-493-19211	Sequence 19211, A
11	749.5	28.1	507	US-10-282-122A-50026	Sequence 50026, A
12	749	28.1	513	US-09-833-745-52	Sequence 52, Appl
13	749	28.1	513	US-09-833-745-58	Sequence 58, Appl
14	748	28.0	522	US-10-369-493-9120	Sequence 9120, Ap
15	738.5	27.7	499	US-10-159-257A-189	Sequence 189, App
16	734.5	27.5	502	US-10-369-493-7044	Sequence 7044, Ap
17	734.5	27.5	511	US-10-282-122A-49086	Sequence 49086, A
18	729.5	27.3	487	US-10-369-493-4288	Sequence 4288, Ap
19	729	27.3	460	US-10-369-493-10228	Sequence 10228, A
20	728.5	27.3	508	US-09-833-745-40	Sequence 40, Appl
21	728.5	27.3	508	US-10-369-493-23365	Sequence 23365, A
22	724	27.1	513	US-09-833-745-46	Sequence 46, Appl
23	721.5	27.0	513	US-09-833-745-44	Sequence 44, Appl
24	717.5	26.9	513	US-09-833-745-50	Sequence 50, Appl
25	715	26.8	494	US-10-369-493-13631	Sequence 13631, A
26	715	26.8	514	US-10-156-761-10862	Sequence 10862, A
27	712	26.7	511	US-10-369-493-17274	Sequence 17274, A

28	711.5	26.7	524	4	US-10-369-493-23452	Sequence 23452, A
29	710.5	26.6	513	3	US-09-833-745-56	Sequence 56, Appl
30	709	26.6	506	4	US-10-369-493-4743	Sequence 4743, Ap
31	709	26.6	506	4	US-10-369-493-7503	Sequence 7503, Ap
32	708.5	26.6	511	4	US-10-282-122A-77245	Sequence 77245, A
33	706	26.5	496	4	US-10-369-493-18118	Sequence 18118, A
34	706	26.5	504	3	US-09-833-745-65	Sequence 65, Appl
35	706	26.5	516	3	US-09-833-745-41	Sequence 41, Appl
36	703.5	26.4	485	4	US-10-369-493-10919	Sequence 10919, A
37	699.5	26.2	484	4	US-10-369-493-8641	Sequence 8641, Ap
38	695	26.0	510	4	US-10-369-493-11500	Sequence 11500, A
39	694	26.0	497	4	US-10-369-493-7583	Sequence 7583, Ap
40	690	25.9	513	3	US-09-833-745-45	Sequence 45, Appl
41	688.5	25.8	511	4	US-10-282-122A-44428	Sequence 44428, A
42	688	25.8	507	4	US-10-369-493-15414	Sequence 15414, A
43	688	25.8	507	4	US-10-369-493-15782	Sequence 15782, A
44	688	25.8	507	4	US-10-369-493-16165	Sequence 16165, A
45	687	25.7	482	4	US-10-369-493-4824	Sequence 4824, Ap

ALIGNMENTS

RESULT 1
US-10-621-826-3
; Sequence 3, Application US/10621826
; Publication No. US20040059103A1
; GENERAL INFORMATION:
; APPLICANT: E.I. du Pont de Nemours and Company, Inc.
; APPLICANT: Huang, Lixuan
; APPLICANT: Xue, Zhixiong
; TITLE OF INVENTION: DNA and Amino Acid Sequences of a Tyrosine Ammonia Lyase Enzyme
; FILE REFERENCE: from the Bacterium Rhodobacter sphaeroides
; CURRENT APPLICATION NUMBER: US/10/621.826
; CURRENT FILING DATE: 2003-07-17
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 526
; TYPE: PRT
; ORGANISM: R. sphaeroides
US-10-621-826-3

Query Match	100.0%;	Score 2668;	DB 4;	Length 526;
Best Local Similarity	100.0%;	Pred. No. 5.5e-214;		
Matches 526;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	VKPMLAMSPKPAVELDRHIDLDQAHAVASGGARIVLAPPARDRCRASEARLGAVIREAR	60	
Db	1	VKPMLAMSPKPAVELDRHIDLDQAHAVASGGARIVLAPPARDRCRASEARLGAVIREAR	60	
Qy	61	HVYGLTTGFGPLANRLISGENVRTLQANLVHHLASGVGVLDTTARAMVLARLVSTAQG	120	
Db	61	HVYGLTTGFGPLANRLISGENVRTLQANLVHHLASGVGVLDTTARAMVLARLVSTAQG	120	
Qy	121	ASGASGTTARLIDLINSELAPVPSRGTVGASGDLTPLAHMVLCLQGRGDFLDRDGTRL	180	
Db	121	ASGASEGTTARLIDLINSELAPVPSRGTVGASGDLTPLAHMVLCLQGRGDFLDRDGTRL	180	
Qy	181	DGAEGLRGRLOPLDLSHRDALVNGTSAVTGIALVNAHACRLHGNWAVALTALLAECL	240	
Db	181	DGAEGLRGRLOPLDLSHRDALVNGTSAVTGIALVNAHACRLHGNWAVALTALLAECL	240	
Qy	241	RGRTEAWAALSDLRPHPGQKDAARLRARVDGSARVVRHVIAERRLDAGDIGTEPEAGQ	300	
Db	241	RGRTEAWAALSDLRPHPGQKDAARLRARVDGSARVVRHVIAERRLDAGDIGTEPEAGQ	300	
Qy	301	DAYSRLCAPQVLGAGFDTLAWHDRVLTIELNAVTDNVPFPDGSVPALHCGNFMGQHV	360	
Db	301	DAYSRLCAPQVLGAGFDTLAWHDRVLTIELNAVTDNVPFPDGSVPALHCGNFMGQHV	360	
Qy	361	TSDALATATVVIAGLAERQIARLTDRLNKGLPFLHRGPAGLNSGFMGAQVTTALLAE	420	

Db 361 TSDALATAVTVLAGLAERQIARLTDERLNRLGPPFLHRGPAGLNSGPMGAQVATALLAE 420
Qy 421 MRATGPASIIHSISTNAANQDVSVLGTIAARLCREKIDRWAEIILALCLAQAAELRCGS 480
Db 421 MRATGPASIIHSISTNAANQDVSVLGTIAARLCREKIDRWAEIILALCLAQAAELRCGS 480
Qy 481 GLDGVSPAGKKLVQALREQPPPLETDRPLGQEIATAALATHLLQQSPV 526
Db 481 GLDGVSPAGKKLVQALREQPPPLETDRPLGQEIATAALATHLLQQSPV 526

RESULT 2

US-11-069-633-8
; Sequence 8, Application US/11069633
; Publication No. US20050208643A1
; GENERAL INFORMATION:
; APPLICANT: Schmidt-Dannert, Claudia
; APPLICANT: Watts, Kevin
; TITLE OF INVENTION: Flavonoids
; FILE REFERENCE: 09531-160001
; CURRENT APPLICATION NUMBER: US/11/069,633
; CURRENT FILING DATE: 2005-03-01
; PRIOR APPLICATION NUMBER: US 60/549,077
; FILING DATE: 2004-03-01
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 526
; TYPE: PRT
; ORGANISM: Rhodobacter sphaeroides
US-11-069-633-8

Query Match 99.9%; Score 2665; DB 6; Length 526;
Best Local Similarity 99.8%; Pred. No. 9.8e-214;
Matches 525; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 1 VKPWLMSPPKPAVELDRHLDLQAHAVASGGARIVLAPPARCRASEARLGAVTREAR 60
Db 1 MKPWLMSPPKPAVELDRHLDLQAHAVASGGARIVLAPPARCRASEARLGAVTREAR 60
Qy 61 HVYGLTTGFGPLANRLISGENVRTLQANLVHLLASGVGPVLDWTARAVLARLVSTIAQG 120
Db 61 HVYGLTTGFGPLANRLISGENVRTLQANLVHLLASGVGPVLDWTARAVLARLVSTIAQG 120
Qy 121 ASGASEGTIARLIDLLNSELAPAVPSRGTVGASGDLTPLAHMVLCLQGRGDFLDRDGTTL 180
Db 121 ASGASEGTIARLIDLLNSELAPAVPSRGTVGASGDLTPLAHMVLCLQGRGDFLDRDGTTL 180
Qy 181 DGAEGLRGRQLQPLDLSHRDALALVNGTSAMTGTALVNAHACRHLGNWVALTALLAECL 240
Db 181 DGAEGLRGRQLQPLDLSHRDALALVNGTSAMTGTALVNAHACRHLGNWVALTALLAECL 240
Qy 241 RGRTEAWAALSDLRPHPGKDAAARLARVDSGARVVRHVIAERRLDAGDIGTEPEAGQ 300
Db 241 RGRTEAWAALSDLRPHPGKDAAARLARVDSGARVVRHVIAERRLDAGDIGTEPEAGQ 300
Qy 301 DAYSLRCAPOVLGAGFTFLAHDRLVTIELNAVTDNVPFPPDGSVPALHGNFMGQHVVAL 360
Db 301 DAYSLRCAPOVLGAGFTFLAHDRLVTIELNAVTDNVPFPPDGSVPALHGNFMGQHVVAL 360
Qy 361 TSDALATAVTVLAGLAERQIARLTDERLNRLGPPFLHRGPAGLNSGPMGAQVATALLAE 420
Db 361 TSDALATAVTVLAGLAERQIARLTDERLNRLGPPFLHRGPAGLNSGPMGAQVATALLAE 420
Qy 421 MRATGPASIIHSISTNAANQDVSVLGTIAARLCREKIDRWAEIILALCLAQAAELRCGS 480
Db 421 MRATGPASIIHSISTNAANQDVSVLGTIAARLCREKIDRWAEIILALCLAQAAELRCGS 480
Qy 481 GLDGVSPAGKKLVQALREQPPPLETDRPLGQEIATAALATHLLQQSPV 526
Db 481 GLDGVSPAGKKLVQALREQPPPLETDRPLGQEIATAALATHLLQQSPV 526

RESULT 3

US-10-369-493-7816
; Sequence 7816, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 7816
; LENGTH: 482
; TYPE: PRT
; ORGANISM: Rhodobacter sphaeroides
US-10-369-493-7816

Query Match 91.6%; Score 2444; DB 4; Length 482;
Best Local Similarity 100.0%; Pred. No. 2.5e-195;
Matches 482; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 43 DRCRASEARLGAIVIREARHVYGLTTGFGPLANRLISGENVRTLQANLVHLLASGVGPVLD 102
Db 1 DRCRASEARLGAIVIREARHVYGLTTGFGPLANRLISGENVRTLQANLVHLLASGVGPVLD 60
Qy 103 WTTARAVLARLVSTIAQASGASEGTIARLIDLLNSELAPAVPSRGTVGASGDLTPLAHM 162
Db 61 WTTARAVLARLVSTIAQASGASEGTIARLIDLLNSELAPAVPSRGTVGASGDLTPLAHM 120
Qy 163 VLCLQGRGDFLDRDGTTLGAGLRGRQLQPLDLSHRDALALVNGTSAMTGTALVNAHAC 222
Db 121 VLCLQGRGDFLDRDGTTLGAGLRGRQLQPLDLSHRDALALVNGTSAMTGTALVNAHAC 180
Qy 223 RHLGNWVALTALLAECLRGRTEAWAALSDLRPHPGKDAAARLARVDSGARVVRHV 282
Db 181 RHLGNWVALTALLAECLRGRTEAWAALSDLRPHPGKDAAARLARVDSGARVVRHV 240
Qy 283 AERRLDAGDIGTEPEAGQDAYSLRCAPOVLGAGFTFLAHDRLVTIELNAVTDNVPFPPD 342
Db 241 AERRLDAGDIGTEPEAGQDAYSLRCAPOVLGAGFTFLAHDRLVTIELNAVTDNVPFPPD 300
Qy 343 GSPALHGNFMGQHVVALTSDALATAVTVLAGLAERQIARLTDERLNRLGPPFLHRGPAG 402
Db 301 GSPALHGNFMGQHVVALTSDALATAVTVLAGLAERQIARLTDERLNRLGPPFLHRGPAG 360
Qy 403 LNSGFMGAQVATALLAEEMRATGPASIIHSISTNAANQDVSVLGTIAARLCREKIDRWAEI 462
Db 361 LNSGFMGAQVATALLAEEMRATGPASIIHSISTNAANQDVSVLGTIAARLCREKIDRWAEI 420
Qy 463 LAILALCLAQAAELRCGSLDGVSPAGKKLVQALREQPPPLETDRPLGQEIATAALATHLLQ 522
Db 421 LAILALCLAQAAELRCGSLDGVSPAGKKLVQALREQPPPLETDRPLGQEIATAALATHLLQ 480
Qy 523 QS 524
Db 481 QS 482

RESULT 4

US-10-464-609-4
; Sequence 4, Application US/10464609
; Publication No. US20040029230A1
; GENERAL INFORMATION:
; APPLICANT: KYNDT, John, Jozef Armand
; APPLICANT: VAN BEEUMEN, Jozef

; TITLE OF INVENTION: No. US20040029230A1el Methods For Synthesis of
; FILE REFERENCE: 50304/008001
; CURRENT APPLICATION NUMBER: US/10/464,609
; PRIOR FILING DATE: 2003-06-18
; PRIOR APPLICATION NUMBER: US 60/389,593
; PRIOR FILING DATE: 2002-06-18
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 540
; TYPE: PRT
; ORGANISM: Rhodobacter capsulatus
US-10-464-609-4

Query Match 50.0%; Score 1335; DB 4; Length 540;
Best Local Similarity 56.5%; Pred. No. 1.3e-102;
Matches 292; Conservative 59; Mismatches 162; Indels 4; Gaps 4;

QY 11 KPAVELDRHIDLDQAHAVASGGARIVLAPPARDRCSEARLGAIVREARHYVGLTTGFG 70
DB 19 KDCALDQALTLVQCEATATHRISVTPALRERCARAHARLEHAIAQRHIYGITGFG 78
QY 71 PLANRLISENVRTLOANLVHILASGVPLDWTTRAMVLARLVSIQAQSGASEGTIA 130
DB 79 PLANRLICADQAELOQNLIIHLATGVGPKLSWAEARALMARLNSIIQAGSASPETID 138
QY 131 RLIDILNSELAPVPSRGTVGASGDLTPLAHMVLCLQGRGDFLDRDTRLDGAEGLRGR 190
DB 139 RIVAVINAGFAPEVPAQTVGASGDLTPLAHMVLALQGRWIDPSGRVQEGAVMDRLC 198
QY 191 LQPLDLHRDALVNGTSAMTGI-ALVNAHACHRLGNWAVALTALLAECLRGTRTEAWAA 249
DB 199 GGPLTLAARDGLALVNGTSAMTATAALTCVEAARAI-DAALRHSAVLMEVLSGHAEAWHP 257
QY 250 ALSDLRPHGQKDAARLARVDSGARVVRHVIARERLDAGDIGTEPEAGODAYSRLCAP 309
DB 258 AFALRPHPGQURATERLAQALDAGRVCRILTAAARLTAAADLRPEHPAQDAYSRLVVP 317
QY 310 QVLGAGFTFLAHDVRLTIELNAVTDNVPFPPDGSVPALHGNFMGQHVALTSDALATAV 369
DB 318 QLVGNVMTLDHWDVVTCELSVTDNPIFEGCAVPALHGNFMGQHVALTSDALNAAL 377
QY 370 TVLAGLAERQIARLTDERLNGLPFLHRGPAAGNSGFMGAQVTTATALLAENRATG-PAS 428
DB 378 VTLAGLVERQIARLTDEKLNGLPAFLHGGQAGLQSGFMGAQVTTATALLAENRATPV 437
QY 429 IHSISTNAANDVVSGLGTIAARLCKEIKDRWAEILAILALCLAQAAELRCG-SGLDGVP 487
DB 438 VQSLSTNGANDVVSVMGTIAARRARAQLPLPSIQIQAIALALALAQAMDLLDPEGQAGWSL 497
QY 488 AGKKLVQALREQFPPLTDRPLGQEIATAALATHLLQSS 524
DB 498 TARDLRDRIRAVSPGLRADRPLAGDIEVAQGLRHPS 534

RESULT 5
US-10-369-493-8812
; Sequence 8812, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 8812
; LENGTH: 529
; TYPE: PRT
; ORGANISM: Ralstonia metallidurans
US-10-369-493-8812

Query Match 31.4%; Score 837; DB 4; Length 529;
Best Local Similarity 40.9%; Pred. No. 5.2e-61;
Matches 210; Conservative 77; Mismatches 209; Indels 18; Gaps 8;

QY 19 HIDLDQAHAVASGGARIVLAPPARDRCSEARLGAIVREARHYVGLTTGFGPLANRLIS 78
DB 9 HLTPTVAAIARGQPAIIVPEVLGKVADARAFQVAAANVPYIGVSTGFGELVHNVWD 68
QY 79 GENVRTLOANLVHILASGVPLDWTTRAMVLARLVSIQAQSGASEGTIARLIDLNS 138
DB 69 IEHGRALQENLRLSHCHACGVPLFSRDEVRAVMVASANALARGYSAPRPAVIEQLKYLEA 128
QY 139 ELAPVPSRGTVGASGDLTPLAHMVLCLQGRGDFLDRDTRLDGAEGLRGRLOPLDL 198
DB 129 GITPAVPQVGSIGASGDLAPLSHVAITLIGEGKVLTDGGTAPTAEVLRERGITFLALAY 188
QY 199 RDALALVNGTSAMTGIALVNAHACHRLGNWAVALTALLAECLRGTRTEAWAAALS 257
DB 189 KEGALALINGTSAMTGVSCILLETTLRAQVQQAETIALLALEGLSADAFMAHGHDIAPKH 248
QY 258 PQQKDAARLARVDSGARVVRH--VIAERRLDAGD---IGTEPEAG---QDAYSRLCAP 309
DB 249 PQQIRSAANMEALLADLSARLSGHGELSDEMTRAGEAKNTGT---GVPIQKAYTLRCIP 304
QY 310 QVLGAGFTFLAHDVRLTIELNAVTDNVPFPPDGSVPALHGNFMGQHVALTSDALATAV 369
DB 305 QVLGAVRDTLDHCAVTVVERELNSNDNPLFFEDGEL--FHGNGHFGQVQVAFAMDFLATAA 362
QY 370 TVLAGLAERQIARLTDERLNGLPFLHRGPAAGNSGFMGAQVTTATALLAENRA-TGRPAS 428
DB 363 TOLGVVSERRNLRLSLPHLNNLPAFLAAANEGLSGFGAGQYPATALLAENRTTICSPAS 422
QY 429 IHSISTNAANDVVSGLGTIAARLCKEIKDRWAEILAILALCLAQAAELRCGSGLDGVS 488
DB 423 IQSVPSNGNDQVVSVMGLIARNARRILDNQYIILALELLASQAAEL--AGAVEQLAPA 480
QY 489 GKLVQALREQFPPLTDRPLGQEIATAALATHLLQ 522
DB 481 GRAVFAFVRERVPFLSIDRYMTDDIEAMALLRQ 514

RESULT 6
US-09-833-745-49
; Sequence 49, Application US/09833745
; Patent No. US20020052038A1
; GENERAL INFORMATION:
; APPLICANT: ROBERTS, JOSEPH
; APPLICANT: SETHURAMAN, NATARAJAN
; APPLICANT: MACALLISTER, THOMAS
; TITLE OF INVENTION: CLONING, OVEREXPRESSION AND THERAPEUTIC USE OF
; FILE REFERENCE: 078728/0106
; CURRENT APPLICATION NUMBER: US/09/833,745
; PRIOR FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/197,770
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 49
; LENGTH: 513
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-833-745-49

Query Match 29.0%; Score 774.5; DB 3; Length 513;
Best Local Similarity 36.4%; Pred. No. 8.4e-56;

Db 185 LQAKEGLALLNGTQASTALALYNMFAIEDLYRTALVSGALSVDAAMGSKPFDARIHEL 244
Qy 256 PHPGQKDAARLARVDSARVVRHVIARERLDAGDIGTEPEAGODAYSILRCAPQVLGAG 315
Db 245 GHQGGIDAAAAYRSLLGSAINVSHADCDK-----VQDPYSLRCQPQVMGAC 291
Qy 316 FDTLAWHDRVLTIELNAVTDNP-VFPDPGSPALHGGNFMGQHVALTSDALATATVVLG 374
Db 292 LDQWRHAANVLLLEANAVIDNPLIFPDTEGEV--LSGGNFHAEPVAFADNLAALAAETGA 349
Qy 375 LAERQIARLTDERLNRLGPPFLHRRGPAGLNSGFMGAQVTTATALLAEMRATG-PASIHIS 433
Db 350 LAERIALLDATLS-GLPFLVR-DGVNSGFMIAHTAALASENKTLAHPASVDSL 407
Qy 434 TNAANQDVSLGTTAARLCREKIDRWAEILAILALCLAAELRCGSLDGVSPAGKLV 493
Db 408 TSANQEDHVSMTAFARKLGDIAENTANILSIELLAAAQGVDLRAP---HKTSPSLQAM 464
Qy 494 QALREQPPLETDRPLQGEIAALATHLLQQSPV 526
Db 465 DTVRKOVAYELDHVFADPAAV-TRLVQNGTI 496

RESULT 9

US-10-282-122A-49837
; Sequence 49837, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 49837
; LENGTH: 507
; TYPE: PRT
; ORGANISM: Burkholderia fungorum
US-10-282-122A-49837

Query Match 28.3%; Score 755; DB 4; Length 507;
Best Local Similarity 38.8%; Pred. No. 3.5e-54;
Matches 199; Conservative 73; Mismatches 211; Indels 30; Gaps 11;
Qy 19 HIDLQAHAVASGGARIVLAP---PARDRCRASEARLCAVIREARHVYGLTTGFGPLANR 75
Db 8 HLTLPQLRQIAREHVALQDPASHAIDACAQAVADIAA---KGEPAVINGTGFGLAST 64
Qy 76 LISGNVTRTQANLVHHLASGVPLDWTTRARVLAARLVIAQAGSASGTIARLIDL 135
Db 65 HIPHDQLELLQNLVLSHAVGVGPMSPVVRLLIALKLSLGRGSHGIRREVMALITL 124
Qy 136 LNSELAAPASRGVVGASGDLTPLAHVVLCLQGRGDFLDRDGTLDGAEGLRRGLQPLD 195
Db 125 YNADVLVPVPGVSGVSGDLPALAHMSAALLGVGEVFAK--GERMPATEGLALVGLKPLT 183
Qy 196 LSHRDALALVNGTSMATGIALVNAHACHRLGNVAVALTALLAECLRGRTAEWAALSDLR 255
Db 184 LQAKEGLALLNGTQASTALALYNMFAIEDLYRTALVSGALSVDAAMGSKPFDARIHEL 243
Qy 256 PHPGQKDAARLARVDSARVVRHVIARERLDAGDIGTEPEAGODAYSILRCAPQVLGAG 315
Db 244 GHQGGIDAAAAYRSLLGSAINVSHADCDK-----VQDPYSLRCQPQVMGAC 290
Qy 316 FDTLAWHDRVLTIELNAVTDNP-VFPDPGSPALHGGNFMGQHVALTSDALATATVVLG 374
Db 291 LDQWRHAANVLLLEANAVIDNPLIFPDTEGEV--LSGGNFHAEPVAFADNLAALAAETGA 348
Qy 375 LAERQIARLTDERLNRLGPPFLHRRGPAGLNSGFMGAQVTTATALLAEMRATG-PASIHIS 433
Db 349 LAERIALLDATLS-GLPFLVR-DGVNSGFMIAHTAALASENKTLAHPASVDSL 406
Qy 434 TNAANQDVSLGTTAARLCREKIDRWAEILAILALCLAAELRCGSLDGVSPAGKLV 493
Db 407 TSANQEDHVSMTAFARKLGDIAENTANILSIELLAAAQGVDLRAP---HKTSPSLQAM 463
Qy 494 QALREQPPLETDRPLQGEIAALATHLLQQSPV 526
Db 464 DTVRKOVAYELDHVFADPAAV-TRLVQNGTI 495

RESULT 10
US-10-369-493-19211
; Sequence 19211, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 19211
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
US-10-369-493-19211

Query Match 28.2%; Score 753.5; DB 4; Length 501;
Best Local Similarity 39.0%; Pred. No. 4.6e-54;
Matches 196; Conservative 81; Mismatches 197; Indels 29; Gaps 9;
Qy 20 IDLDOAHAVASGGARIVLAPPARDRCRASEARLCAVIREARHVYGLTTGFGPLANRLISG 79
Db 8 LKUEILQVARNATVELSPDAATVRASRALVDRVAAGDTPAYGINTGFTLAEVRIDK 67

Query Match 28.1%; Score 749; DB 3; Length 513;
Best Local Similarity 37.4%; Pred. No. 1.1e-53;
Matches 199; Conservative 91; Mismatches 194; Indels 48; Gaps 15;

QY 12 PAVELD--RHIDLDQAHVAGGARGIVLAPPADRCRASEARLGA---VIREARHVYGLTT 67
Db 5 PMIEIDGRSLRVEDVYAVAYEYDVSTSD---DTLKAAVEEKHEAFKLKINSCKTVGYNT 61

RESULT 14
US-10-369-493-9120
; Sequence 9120, Application US/10369493
; Publication No. US2003023675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF PHARMACEUTICALS
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 9120
; LENGTH: 522
; TYPE: PRT
; ORGANISM: Chloroflexus aurantiacus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(522)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-9120

Query Match	28.0%;	Score 748;	DB 4;	Length 522;
Best Local Similarity	39.0%;	Pred. No. 1.4e-53;		
Matches 205;	Conservative	70;	Mismatches 205;	Indels 46;
				Gaps 13

[illegible]

```

RESULT 15
US-10-159-257A-189
; Sequence 189, Application US/10159257A
; Publication No. US20040161828A1
; GENERAL INFORMATION:
; APPLICANT: SHEN, BEN
; APPLICANT: LIU, WEN
; APPLICANT: CHRISTENSON, STEVEN D.
; APPLICANT: STANDAGE, SCOTT
; TITLE OF INVENTION: GENE CLUSTER FOR PRODUCTION OF THE ENEDIYNE ANTITUMOR
; TITLE OF INVENTION: ANTIBIOTIC C-1027
; FILE REFERENCE: 407T-896020US
; CURRENT APPLICATION NUMBER: US/10/159,257A
; CURRENT FILING DATE: 2002-05-31
; PRIOR APPLICATION NUMBER: 09/478,188
; PRIOR FILING DATE: 2000-01-05
; PRIOR APPLICATION NUMBER: 60/115,434
; PRIOR FILING DATE: 1999-01-06
; NUMBER OF SEQ ID NOS: 207
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 189
; LENGTH: 499
; TYPE: PRT
; ORGANISM: Streptomyces globisporus
; FEATURE:
; OTHER INFORMATION: orf24
US-10-159-257A-189

```

	Query Match	27.7%;	Score 738.5;	DB 4;	Length 499;
	Best Local Similarity	42.6%;	Pred. No. 8.2e-53;		
	Matches 182;	Conservative 59;	Mismatches 163;	Indels 23;	Gaps 7;
Qy	106	ARAWLARLVSYAAGSAGSEGTIARLIDLINSELAPVSRGVTWGASGDLTPLAHVLC	165		
Db	66	ARAIVAARLNTLAKGHSAVRPIILERLAQYINEGITPAIPGISAGSGDLAPLUSHVAST	125		
Qy	166	LQGRGDFLDGRTLDCAEGLRRRLQPLDLSHRDALVNCTGTSAMTGIALVNAHACRHL	225		
Db	126	LIGEG-VYLRGRPVETAQVLAERGIEPLELRFKEGLALNGTSGMTGLGSLVVGRALEQ	184		
Qy	226	GNNAVALTALLAECLRGRTBWAALSDL-RPHFGQKDAARLARLRVDGSAVRVHVIAE	284		
Db	185	AQAEIVTALLIEAVRGSTSPFLAEGHDIAAPHEGOIDTRANMMLMRGSLTVEHADLR	244		

Qy	285	RRLDAGDIGTEPRAGQD-----AYSILRCAPOVLGAGFDTLAWHDRVLTIELNAV	334
Db	245	RELQ-----KDEKAGKDVQRSEIYLQKAYSLRAIPQVVGAVRDTLYHARHKLR	299
Qy	335	DNPVFPDGSGVPALHGGNFMGQHVALTSDALATAVTVLAGLAERQIARLTDERINRGL	394
Db	300	DNLPFPPEGKEI--FHGANFHGQPIAFAMDFTVIALTQVLAEQINRVNLRHLSYGL	357
Qy	395	FLHRGPAGLNSGFMGAQVTTALLAENRATGPASIHISITNNAQDVVVSGLGTIAAR	454
Db	358	FLVSGDGFLHSGFAGAQYPATALVAENRTITGPASTQSPVSGNDQDVVSMGLIS	417
Qy	455	KIDRWAEILAILALCLAAOAEELRCGSG-LDGVSVPAGKKLVQALREQPPPLETDR	513
Db	418	VLNNKKILAVEYLAAAQAVDI---SGRFGLSPAAKATYEAVRRLVPTLGDVRYM	474
Qy	514	AALATHL 520	
Db	475	ELVADAL 481	

Search completed: December 4, 2005, 09:19:12
Job time : 167 secs

Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	461	17.3	716	7	US-11-105-259-2	Sequence 2, Appli	
2	104.5	3.9	470	6	US-10-979-821-14	Sequence 14, Appl	
3	103	3.9	566	6	US-10-467-657-3302	Sequence 3302, Ap	
4	101.5	3.8	1124	6	US-10-858-730-12	Sequence 12, Appl	
5	100	3.7	397	6	US-10-467-657-9206	Sequence 9206, Ap	
6	100	3.7	417	6	US-10-467-657-7708	Sequence 7708, Ap	
7	99.5	3.7	1694	7	US-11-135-855-36	Sequence 36, Appl	
8	99.5	3.7	1709	7	US-11-135-855-35	Sequence 35, Appl	
9	98	3.7	628	7	US-11-147-047-45	Sequence 45, Appl	
10	95.5	3.6	2004	6	US-10-467-657-84	Sequence 84, Appl	
11	95.5	3.6	2004	6	US-10-467-657-6322	Sequence 6322, Ap	
12	93	3.5	479	7	US-11-074-176-168	Sequence 168, App	
13	93	3.5	479	7	US-11-147-047-44	Sequence 44, Appl	
14	93	3.5	553	6	US-10-821-234-1636	Sequence 1636, Ap	
15	93	3.5	776	6	US-10-925-970-3	Sequence 3, Appl	
16	92	3.4	1476	6	US-10-647-956A-4	Sequence 4, Appl	
17	91	3.4	511	6	US-10-467-657-4656	Sequence 4656, Ap	
18	90	3.4	514	6	US-10-467-657-2864	Sequence 2664, Ap	
19	90	3.4	768	7	US-11-186-731-5	Sequence 5, Appl	
20	89.5	3.4	777	6	US-10-467-657-2474	Sequence 2474, Ap	
21	89	3.3	557	6	US-10-512-109-9	Sequence 9, Appl	
22	88.5	3.3	396	7	US-11-061-869-11	Sequence 11, Appl	
23	88.5	3.3	417	6	US-10-467-657-8450	Sequence 8450, Ap	
24	88.5	3.3	934	6	US-10-858-730-8	Sequence 8, Appl	
25	87.5	3.3	357	7	US-11-055-822-734	Sequence 734, App	

Db 339 SRFAVHH---EEVVKVD--DEGILRQDRYPLRTSPQWLGPVSDLIHAHVITIEGQS 393
QY 333 VTNPVPPDGSPALHGGFMGQHVATSDALATAVTVVLGALAEQIARLTDERLNRGL 392
Db 394 TTDNPLDVENKT-SHEGGNFQAAAVANTWETKTRGLAQIKLNFQLTETMLNAGMNRGL 452
QY 393 PPFLHRGPAGINSFGMAQVATATALLAEM-RATGPASIHISINNAQDVVSIGTTIARL 451
Db 453 PSCLAABDPSPSYHCKGLDIAAAAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARR 512
QY 452 CREKIDRWAETLALCLQAABELRGSGLDGVSAGKGLVQALREOPFPPELTDRLPGQ 511
Db 513 TTESNDVLSILLATHLYCVLQADLR-AIEFEFKQFGPAIVSLIDQFGSAMTGSNLRD 571
QY 512 EIA-----ALATHLLQ 522
Db 572 ELVEKNKNTLAKRLEQ 587

RESULT 2
US-10-979-821-14
; Sequence 14, Application US/10979821
; Publication No. US20050244937A1
; GENERAL INFORMATION:
; APPLICANT: ABRAHAM, TIMOTHY W.
; APPLICANT: CAMERON, DOUGLAS C.
; APPLICANT: HICKS, PAULA M.
; APPLICANT: MCFARLAN, SARA C.
; APPLICANT: MILLIS, JIM
; APPLICANT: ROSAZZA, JACK
; APPLICANT: ZHAO, LISHAN
; TITLE OF INVENTION: POLYPEPTIDES AND BIOSYNTHETIC PATHWAYS FOR THE
; FILE REFERENCE: 023829-0390
; CURRENT APPLICATION NUMBER: US/10/979,821
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: 10/422,366
; PRIOR FILING DATE: 2003-04-23
; PRIOR APPLICATION NUMBER: 60/374,831
; PRIOR FILING DATE: 2002-04-23
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: Patentin ver. 3.3
; SEQ ID NO 14
; LENGTH: 470
; TYPE: PRT
; ORGANISM: Rhodobacter sphaeroides
US-10-979-821-14

Query Match 3.9%; Score 104.5; DB 6; Length 470;
Best Local Similarity 25.2%; Pred. No. 0.29;
Matches 117; Conservative 31; Mismatches 152; Indels 165; Gaps 22;

QY 1 VKPM-----LAMPFPKPAVEL--DRHIDLQAHAVASGGARIVLAPARDRCRASEARLG 53
Db 98 VEPVAMRAGLAFSDGAPDPELPDK-----ALAFARFALLSPAFRAGADYGDARGT 149
QY 54 AVTIREARHVYGLTTGFGFLANRLISGENVRLQANLVHHLASGVGVPLDWTARTAMVLR 113
Db 150 SSUREA-----LAAYLASDRGVVAD-----PAR 172
QY 114 LVSIAGQASGASEGTIARLIDLINSELAP----AVPSRGTVGASGDILPLAHMVLCLQG- 168
Db 173 LL-LARGSQMA-----LFLVARAALPGEATAVEEPG-----YPLAWEAFRAAGA 216
QY 169 --RGDFLDRDGTLDGAE-----GLRRGLQPLDLSHRD 200
Db 217 EVRGVPPVDGGLRLDALEAALARDPRIRAVYVTPHHQYPTVTWGAAR-RLQLELAERH 275
QY 201 ALALVN-----GTSAMTGIALVNAHACRHLGNWAVAL-TALLAECLRG 242
Db 276 RLALIEDDYDHYRFEGRPVPLAARAPEGPLIYVGSLSKLLSPGIRUGYALAPERLLT 335

QY 243 RTAWAAALSDLRPHFGQKDAARLARVDGS-----ARVVRHVIAERRLD----- 288
Db 336 RMAAARAAIDROGDAFLEAALAEIR---DGDIGRHAKARVYRARR-DLLAERLTAQL 391
QY 289 AGDIGTEPEAGQDAYSLRCAPQVLGAGFDTLAWHDRVLTIELNAVTDNVPFPPDGSPAL 348
Db 392 AGRAADFLPAGGLALWLR-----AGVSAETWAEAA-----CGAGLALL 430
QY 349 HGGFMGQHVATSDALATAVTVVLGALAEQIARLTDERLNRGLP 393
Db 431 PGTRF-----ALESAPAFRLGYAALDEGQIARAV-EILARSFP 469

RESULT 3
US-10-467-657-3302
; Sequence 3302, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: Seqwin99, version 1.04
; SEQ ID NO 3302
; LENGTH: 566
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-3302

Query Match 3.9%; Score 103; DB 6; Length 566;
Best Local Similarity 20.6%; Pred. No. 0.48;
Matches 116; Conservative 56; Mismatches 185; Indels 206; Gaps 24;

QY 21 DLDQAHAVASG-----CARIVLAPPARDRCSEARLGAIVREARHVYGLTTGFGPLA 73
Db 82 DADGATACAVGLDGLAAMGAKVDLVPNR-----FEHGYGLTPELAEIA 125
QY 74 -----NRLISGEN-----VRTQANLV---HHL----- 93
Db 126 AAQGVLLITVDNGIASIAGVARAQAQLGDIVTDHHLPADTVPCIIIVNPNGKCGFPS 185
QY 94 --ASGVGPVLDWTTARAVLARLVSTAGASGEGTIARLIDLINSELAPAVPSRGTVG 151
Db 186 KSLAGVGVIYVLTALRAELRRNRYE---SDGIKEPNLGGLLDLV-----ALG 230
QY 152 ASGDLTPLAHMVLCLQGRGDFLDRDTRLDGAEGL-----RRG--RLQPLDLSHRDALAV 205
Db 231 TVADVVPVLDHNNRIILVSQGLKMRSGKMRPGIRALFEVARDWRKQAQPFDM----- 281
QY 206 NGTSAMTGIAL-VNAHACRHLGNWAVALTALLAECLRGTEAWAAALSRLRPHPGOKDAA 264
Db 282 -----GFLGPRINAAGRLDMSVGIACLIA---RDDSEA-----QELA 317
QY 265 ARLRARVDGSARVVRHVIAERLIDAGDIGTEPEAGQDAYSLRCAPQVLGAGFDTLA---- 320
Db 318 ARL-----NNLNIERR-----ETEQSMLRDALNAFPETLPSGQTTLVAYRD 358
QY 321 -WHDRVLTTELNAVTDNVPFPPDGSPALHGGFMGQHVATSDALATAVTVLAGLAERQ 379
Db 359 DFHQGVVGVIVASRLKDRFYRPTIVFAPA-DNGEVRGSGRSIPNLHLRDALDLVSKRHPDL 417
QY 380 IARLTDERLNRGLPPFLHRGPAGLNSGFGMAQVATATALLAEMRATGPASIHISITNAANQ 439
Db 418 ILKFGGHANAAGLSILEHNIPA-----FQTAEEAVREM-----VCEDDLQ 459

Qy 440 DVVSLGTIAARLCKRIDRWAEILAILALCLAAELRC---GSLDGVSPAGKLVQAL 496
Db 460 TYITDGS LPA-----CDITLQEQNLACHVWGQGF----- 489
Qy 497 REQFPFPLETD-----RPLQGE 512
Db 490 ---APPSFTDFHVVRQQLGAE 509
RESULT 4
US-10-858-730-12
; Sequence 12, Application US/10858730
; Publication No. US20050255568A1
; GENERAL INFORMATION:
; APPLICANT: Bailey, Richard B.
; APPLICANT: Blomquist, Paul
; APPLICANT: Doten, Reed
; APPLICANT: Driggers, Edward M.
; APPLICANT: Madden, Kevin T.
; APPLICANT: O'Leary, Jessica
; APPLICANT: O'Toole, George
; APPLICANT: Trueheart, Joshua
; APPLICANT: Walbridge, Michael J.
; APPLICANT: Yorgey, Peter S.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR AMINO ACID
; TITLE OF INVENTION: PRODUCTION
; FILE REFERENCE: 14184-030001
; CURRENT APPLICATION NUMBER: US/10/858,730
; CURRENT FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/475,000
; PRIOR FILING DATE: 2003-05-30
; PRIOR APPLICATION NUMBER: US 60/551,860
; PRIOR FILING DATE: 2004-03-10
; NUMBER OF SEQ ID NOS: 364
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 1124
; TYPE: PRT
; ORGANISM: Streptomyces coelicolor
US-10-858-730-12
Query Match 3.8%; Score 101.5; DB 6; Length 1124;
Best Local Similarity 21.9%; Pred. No. 1.6; Mismatches 226; Indels 203; Gaps 33;
Matches 137; Conservative 61;
Qy 3 PMLAMSPKPAV-ELDRHIDLQ-----AHAVASGGARIV--LAPPARDRCRASEARLGA 54
Db 130 PVLGSSAPSTDVDELVRAD-DVGFPPVVKAVAGGGGGRVVEEPAQLR-----EA-IEA 183
Qy 55 VIRARHYGLTTGFGPLANRLISGENVRTLQANLVHHLASGVGVLDWTTTARAMVLARL 114
Db 184 ASREAASAFGDSVTF--LEKAVVEPRHIE-----VQILDGEGDVHLHFERDCSVQRRH 235
Qy 115 VSIAQAGSASEGVIARLIDLNLSELAPVPSRGTGVSAGDGLTPLAHMVLCLQ----- 167
Db 236 QKVI-----ELAPA-----PMLDPALEKRICADAVNFAQ 265
Qy 168 -----GRGDPL-DRDGTGL-----DGABGLRRGRQLPDLISHRDALALVNGTSAMTG 213
Db 266 IGYENAGTVEFLVDRDGNHVFENMPRIQVETHTVEETVDVLVQSQRITAAAGTLDLGL 325
Qy 214 IALVNAHACHRLGNWVALTALLAECLRGRTTEAWAALSRLRPHPGQKDAARLARVDG 273
Db 326 LAQEN-----ITLGAALQC-RITTEDPA---NGRPDTGQISA-----YRSPG 365
Qy 274 SARVVRHVIARRLDAG--DIGTPEAGODAYSRLCAPQVLGAGFTLAWHDRVLTIE-- 329
Db 366 GSGI-----RLDGGTTAGTETISAHFDSMLVKLSR--GRDFTTAVNARRAVAEFR 415
Qy 330 -----LNATVNDPVPDPDGSVPALHGGNFMGO--HVALTSDALATATVTLAGLAE 377
Db 416 IRGVATNIPFLQAVDDDFDQA-GRVTT-----SFIEORPHLLTARHSADRGTKLLTYLAD 470

Qy 378 QRIARLUTDER-----LNRGLPPFLHR-----GPAGL-----NSGFMAQVTA 414
Db 471 VTNKPHGERPELVDTLKLPTASAGEPPAGSROLLAELGEGEPFARRRESTTIG--VTD 528
Qy 415 TALLAEMRATGPASISHSISTNAANQDVVSLGTIAARLCREKI----- 456
Db 529 TTF-----RDAHQSLLATFVRTKMLAVAPVVRTLPQLLSLECWGATYDVVALRELA 581
Qy 457 -DRWAEILAILA-----LCLAQAELRCGSLDGVSPAGKLVQ-- 495
Db 582 EDPWERLAALREAVPNLCQM---LLRGRNTVGYTPYPTVTEVDAFVQEAATGIDIFRIF 638
Qy 496 -----LREQFPFPLETDRLPQGEIATA 517
Db 639 DALNDVEQMRPAIEAVRQTGSAVEA 665
RESULT 5
US-10-467-657-9206
; Sequence 9206, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 9206
; LENGTH: 397
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-9206
Query Match 3.7%; Score 100; DB 6; Length 397;
Best Local Similarity 22.1%; Pred. No. 0.51; Mismatches 55; Indels 90; Gaps 17;
Matches 78; Conservative 55;
Qy 70 GPLANRLISGENVRTLQANLVHHLASGVGVLDWTTTARAMVLARL-----VSIAQAGSAS 125
Db 81 GTVNNTCAPYELVYRTWEASIL-----VLGP-----TLARFGEAQVSLPGGCAIGS 125
Qy 126 EGTIARL--IDLINSELAPVPSRGTGVSAGDGL--TPLAHMVLCLQGRGDFLDRDGTGLD 181
Db 126 RPVDQHLKGLEAMGAEI---VIEHGYVVKAGKLGKTRVMDVVTVGGTENLL-MAATLAE 181
Qy 182 GABGLRRGRQLPDLISHRDALALVN-----GTSAMTGIALVNAHACHRLGNWVALTAL 235
Db 182 GTTVLENCALEPEVVDLAECLVFMGAKISIGISTMTLVEGVDELHGCHEH-----SV 232
Qy 236 LAECLRGRTTEAWAALSRLRPHPGQKDAARLARVDGSARVVRHVHTAERRLDAGDIGTE 295
Db 233 VPDRIEAGTFLCAVIT-----GGRVVL--NAAPKTMEVVLDKLVEAGAV--- 276
Qy 296 PEAGQD--AVSLRCAPO-----VLGAGFTD-----LAWHDRV 326
Db 277 IEAGDDWIAIDMRQRPKAVDIRTVVHGFPTDMQAFMALNAVAEGSCRVVETTFENRFM 336
Qy 327 TI-ELNAVTDNVPDPDGSVPALHGGNFMGOHVALTSDALATATVTLAGLAE 378
Db 337 HVPENLRMGAN--ITTEGNTAFVQGVVERLSGAVVVKATDLRASASLVITAGLAAR 387
RESULT 6
US-10-467-657-7708
; Sequence 7708, Application US/10467657

```
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 7708
; LENGTH: 417
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-7708

Query Match      3.7%; Score 100; DB 6; Length 417;
Best Local Similarity 22.1%; Pred. No. 0.55;
Matches 78; Conservative 55; Mismatches 130; Indels 90; Gaps 17;

QY 70 GPLANRLISGENVRTLQANLVHLSAGVPLDWTARTAMVLAARL-----VSIAQAGSAS 125
DB 77 GTVNNTCAPVELVTRWASIL-----VLGP-----TLARFGEAQSLSLPGGCAIGS 121

QY 126 EGTIARL--IDLNSELAPVPSRGTVGASGDL--TFLAHMVLCLQGRGDFLDRDGTRLD 181
DB 122 RPVDQHLKLGLEAMGAEI---VIEHGYVYKAGKLGKTRVAMDVVTVGGTENLL--MAATLAE 177

QY 182 GAEGLRGRLOPLDLSHRDALVN-----GTSAMTGIALVNAHACHRLGNWAVALTAL 235
DB 178 GTTVLENCAIEPEVVDLAELCVMGAKISGIGTSTMIVEGVDELHGEH-----SV 228

QY 236 LAELCRGRTEAWAAALSDLRPHPGQDMAARLARVDGSAVRVHVIAERRLDAGDIGTE 295
DB 229 VPRIEAGTFLCAVIT-----GRRVLR--NAAPKTMEVVDKLVEAGAV--- 272

QY 296 PEAGQD--AVSLRCAPO-----VLGAGFDT-----LAMHDRVLL 326
DB 273 IEAGDDWIAIDMRQRPKAVDVRTVWHPGPTDMQAQFMALNAVAEGSCRVTIFENRFM 332

QY 327 TI-ELNAVTDNPVPPDGSVPALHGGNFMGQHVALTSDALATVTVLAGLAER 378
DB 333 HYPELNRMGAN--ITTEGNTAFVQGVRLSGAVVKATDURASASLVLAGLAAR 383

RESULT 7
US-11-135-855-36
; Sequence 36, Application US/11135855
; Publication No. US20050255557A1
; GENERAL INFORMATION:
; APPLICANT: SMITHKLINE BEECHAM CORPORATION
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP50013
; CURRENT APPLICATION NUMBER: US/11/135,855
; CURRENT FILING DATE: 2005-05-24
; PRIOR APPLICATION NUMBER: US/10/203,708
; PRIOR FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: PCT/US01/04703
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: 60/182,172
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: 60/186,084
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 36
; LENGTH: 1694
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-135-855-35

Query Match      3.7%; Score 99.5; DB 7; Length 1709;
Best Local Similarity 23.1%; Pred. No. 3.8;
Matches 123; Conservative 41; Mismatches 215; Indels 153; Gaps 24;

QY 48 SEARLGAIVREARHVVYGLTTGFGPLANRL--TSGENVRTLOANLVHHLASG----- 96
DB 431 SEPLATLVLSHGHIILASTSGSDSDHSPFRSGTSGPNSLRLETRDLRETSGEYKCSATNS 490

QY 97 ---VGPVLDW--TTARAMVLARLVIAAQG-----ASGASEGTIARLIDLIN----- 137
DB 491 LGNATSTLDFHANAAARLLISPAAEVVEGQAVTLCSCFSGLSPTDFARFSWYNGALLHEGP 550

QY 138 ---SELAPAVPS-----RGTVGASGDLTPLAHMVLCLOGRGDFLDRDGTRLD----- 181
DB 551 GSSLLLPAAASSTDAGSYHCRARDGHSASGSPSPAVLTVLVYPPRQPTFT-----TRLDLDA 606

QY 182 GAEGLRGRGL-----QPLDLSHRD---ALALVNGTSAWT-----GIA 215
DB 607 GAGARRGLLRCVSDPPARLQLLHKDRVATSLSPSGGGGCTCGGCSPRMKVTKAPNLL 666

QY 216 LVNAH-----ACRHLGNWAVALTALAECLRGRTTEAWAAALSDLRPHPGQKD 262
DB 667 RVEIHNPILLEEGLYLCEASNALGNASTSAT-----FNGQATVLAAPS-----HTLQEG 716

QY 263 AARLARVARVDGSA-----RVVRHVIAERRLDAGDIGTEPEAGQDAYSLRCAPOVLGAG 315
DB 717 TEANLTCNVYSREAAGSPANFWSFRNGVLWAQGFLETVTLLPVARTDAALYAC----- 768

QY 316 FDTLAWHDRVLTIELNAVTDNPV-----PDPDGSVPALHGGNFMGO--HVAL---TSDALA 366
DB 769 -----RILT-EAGAQLSTVLLSVLYPPDR--PKLSALLDMGGCHWALFCTVDSRP 817

QY 367 TAVTVLAGLAERQIARLTDRLNRLGPLPFLHRGPAGLNSGFMGAQVTTATALLAEMRATGP 426
DB 818 LALLALFH-GEHLA-----TSLGPQVPSHGRFOAKAANSKLKLEVEGL 862

QY 427 ASHSISTVNAQDVVSLOTIARLCREKIDRWAEITLAILALCLAQAAELRC 478
DB 863 GDSGSYRCATN-----VLGSSNTSLFFQVRGAVVQVSPSPQLQEGQAVVLS 910

RESULT 8
US-11-135-855-35
; Sequence 35, Application US/11135855
; Publication No. US20050255557A1
; GENERAL INFORMATION:
; APPLICANT: SMITHKLINE BEECHAM CORPORATION
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP50013
; CURRENT APPLICATION NUMBER: US/11/135,855
; CURRENT FILING DATE: 2005-05-24
; PRIOR APPLICATION NUMBER: US/10/203,708
; PRIOR FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: PCT/US01/04703
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: 60/182,172
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: 60/186,084
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 35
; LENGTH: 1709
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-135-855-35

Query Match      3.7%; Score 99.5; DB 7; Length 1709;
```


QY 25 AHAVASGARIVLAPPARDCRSEARLGAVIREARHVYGLTTGFGPLANRLISGENVRT 84
DB 341 SEAVVGGRLTALPTEDKGLAQDVR-----QDVQ--GLTGGGRLTPDAGADANAAA 393
QY 85 LQANLVHHLASGVGVLD---WTTARAVLRLVLSIAQGSASGEGTIA-----RLI 133
DB 394 LQGLPGSAVASGNAPARRQNQLQVRAEAGAAPGLSASENLAGTDGCKRAPVACKRPDTVL 453
QY 134 DLINSELAPVPSRGTVGASGDLTPLAHMVLCLQGRGDFL-----DRDGTLDGA 183
DB 454 PVLNPOVAE-----SAGRVSPKKRMA---DAAADFTRLAADRRRPEKAGVPLGGG 501
QY 184 EGLRRGRLOPLDLSHRDALALVNGTSAMTGI-----ALVNAHACRHL-GNW 228
DB 502 E-----YRFHTDRRHIDALAGVPGKGMPEFADMGAPNSDGLV-SDGRRYLKGRE 556
QY 229 AVALTA-LIAECL-----RGRTEMAAALSDLRPHPGQKDAARLRAR-----V 271
DB 557 AETLRAGGLSEAVPSEPRGRDYPRTQEARAPAKVMARPRDAAADGKPAQAQPARAKDTPV 616
QY 272 DGSARVVHVIAE-----RLDAG-----DIGT-EPEAGQDAYSRLCAPOVLGAGFDTL 319
DB 617 AGKAAAANAATEKPSDKVRIEAGKSRFDGKGKSAAGAAATEKPSKTKAKPETF 676
QY 320 AWHDRVLTIELNAVTDPNVPFPDGSVPALHGG-----NFMGQHVALTSDAL 365
DB 677 A-----KTASDNPE-EARRKARVLQGGPVYTVTKERQAPQGPKALREHAESIKKRL 725
QY 366 ATAVTVLAGLAER 378
DB 726 AESI---GGLAER 735

RESULT 11

US-10-467-657-6322
; Sequence 6322, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON Spa
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 6322
; LENGTH: 2004
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-6322

Query Match 3.6%; Score 95.5; DB 6; Length 2004;
Best Local Similarity 23.3%; Pred. No. 9.7;
Matches 101; Conservative 45; Mismatches 170; Indels 117; Gaps 21;

QY 25 AHAVASGARIVLAPPARDCRSEARLGAVIREARHVYGLTTGFGPLANRLISGENVRT 84
DB 341 SEAVVGGRLTALPTEDKGLAQDVR-----QDVQ--GLTGGGRLTPDAGADANAAA 393
QY 85 LQANLVHHLASGVGVLD---WTTARAVLRLVLSIAQGSASGEGTIA-----RLI 133
DB 394 LQGLPGSAVASGNAPARRQNQLQVRAEAGAAPGLSASENLAGTDGCKRAPVACKRPDTVL 453
QY 134 DLINSELAPVPSRGTVGASGDLTPLAHMVLCLQGRGDFL-----DRDGTLDGA 183
DB 454 PVLNPOVAE-----SAGRVSPKKRMA---DAAADFTRLAADRRRPEKAGVPLGGG 501

QY 184 EGLRRGRLOPLDLSHRDALALVNGTSAMTGI-----ALVNAHACRHL-GNW 228
DB 502 E-----YRFHTDRRHIDALAGVPGKGMPEFADMGAPNSDGLV-SDGRRYLKGRE 556
QY 229 AVALTA-LIAECL-----RGRTEMAAALSDLRPHPGQKDAARLRAR-----V 271
DB 557 AETLRAGGLSEAVPSEPRGRDYPRTQEARAPAKVMARPRDAAADGKPAQAQPARAKDTPV 616
QY 272 DGSARVVHVIAE-----RLDAG-----DIGT-EPEAGQDAYSRLCAPOVLGAGFDTL 319
DB 617 AGKAAAANAATEKPSDKVRIEAGKSRFDGKGKSAAGAAATEKPSKTKAKPETF 676
QY 320 AWHDRVLTIELNAVTDPNVPFPDGSVPALHGG-----NFMGQHVALTSDAL 365
DB 677 A-----KTASDNPE-EARRKARVLQGGPVYTVTKERQAPQGPKALREHAESIKKRL 725
QY 366 ATAVTVLAGLAER 378
DB 726 AESI---GGLAER 735

RESULT 12

US-11-074-176-168
; Sequence 168, Application US/11074176
; Publication No. US20050250135A1
; GENERAL INFORMATION:
; APPLICANT: Kjaenhammer, Todd R.
; APPLICANT: Russell, William M.
; APPLICANT: Altermann, Eric
; APPLICANT: McAuliffe, Olivia
; APPLICANT: Perill, Andrea Azcarate
; TITLE OF INVENTION: Nucleic Acid Sequences Encoding
; TITLE OF INVENTION: Stress-Related Proteins and Uses Therefore
; FILE REFERENCE: 5051-694
; CURRENT APPLICATION NUMBER: US/11/074,176
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: 60/551,161
; PRIOR FILING DATE: 2004-03-08
; NUMBER OF SEQ ID NOS: 381
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 168
; LENGTH: 479
; TYPE: PRT
; ORGANISM: Lactobacillus acidophilus
US-11-074-176-168

Query Match 3.5%; Score 93; DB 7; Length 479;
Best Local Similarity 22.1%; Pred. No. 2.3;
Matches 62; Conservative 36; Mismatches 104; Indels 78; Gaps 11;

QY 271 VDGSARVVV---HVIARRRLDAGDIGTEPEAGQDAYSRL-----CAPOVLGAGFD 318
DB 27 INNALRVKSEDESIVLEVTLELGDGLVLTIAWESTDGLRRGMKVEDTGAPISVPVGEDT 86
QY 319 LAWHDRVLTIELNAVTDPNVPFPDGSVPALHGGNFMGQHVALTSDALATATVTVLAGLAER 378
DB 87 LG---RVFNVLGQPIDGGPAFPKDHPRHIGIHKPEAKPYEDLTTSREILETGKVI----- 137
QY 379 QIARLTDERLNRGLPPLHRGPAGLNSGFMGQAVTATALLAEMRATGPASIHISSTNAAN 438
DB 138 -----DLLEPYVRGGKVGL---FGGAGVGKTTIIQEL-----IHNI---AQE 173
QY 439 QDVVSLGTIAARLCREKIDRWAE-----ILAILALCLAAQAAELRCGSLDGVSPAGKLV 493
DB 174 HGISVFTGVGRTRENDLIYEMKASGVLSKTAMVFGQWNE-----PPGARMR 222
QY 494 QA-----LREQPPPLETDRPL-----GQEIAAL 516
DB 223 VALTGUTLAEIFRDEVBQDVLFIIDNIPRFTQAGSEVSAL 262

RESULT 13

Query Match	3.5%;	Score 93;	DB 7;	Length 479;
Best Local Similarity	24.5%;	Pred. No. 2.3;		
Matches 123;	Conservative 42;	Mismatches 176;	Indels 162;	Gaps 30;
Qy	27	AVASGGARIIVLAPPARDRCASEARLG----	AVIREARHVYGLTTGFGFLANRLISGENV	82
Db	42	SVLCPGAGLIFVPPSLDR--RAEELRADNFASVRR-----	DLAN--MTGLLH	87
Qy	83	RTIQANLVHHLASGVGVLDWTMTARAMVL--ARLVSIQAQ-----		120
Db	88	LSLSRNTIRHVAAGA--FADLRALRALHLDGNRLTSLGEGQLRGLVNLRLHILSNNOQLAA	145	
Qy	121	ASCASGTTARLIDL-----LNSLAP--AVPSGGTVGASGDLTFLAHVWLCLOGRDP-	172	
Db	146	LAAGALDDCAETLEDLDLSYNNLEQLPWEALGRLGNNVTLG-----LDHNLLASVPAGAFS	201	
Qy	173	-----LDRDGTGLDGAQ-----LRRGRLLQPLDLSHRDALALV---NGTSAWMTG	213	
Db	202	RLHKLARLDMTSNRLTTPPPDLESRLPLLARPGSP-----ASALVLAFGNPLHCN	254	
Qy	214	IALV-----NAHACRH---LGN---WAVALTALLAE---CLURGRTEAWAALSRLRP	256	
Db	255	CELVWLRLAREDDLEACASPPALGGRYFWAGVEEFVCEPPVVTHTSRPPLAV-----	307	
Qy	257	HPGQKAAARLARVDGSARVVRHVIARERLDAGDIGTEPPAGODAYSLRCAQVQLGAGF	316	
Db	308	PAGRPAAALRCRAVGDPPEPR--VRWVSPQGRL-----LGNSSRA-----RAPFNG-----	349	
Qy	317	DTLAWHDRVLTIELNAVTDNPVPPDGSVPALHGGNFMGQHVALTSDALATAVTVLAG-L	375	
Db	350	-----TLEL-LVTE-----PDDGGIFTICIAANRAGEATA-----AVELTVGPP	386	
Qy	376	AERQIARLTDERLNRGLPPFLHRRGPAGLNSGF-----MGAQVTV---ATALLA---	419	
Db	387	PPQLANSTCDPPRCDGDPDALTPPSAASAKVADTGPPPTDRGVQVTEHGATAALVQWP	446	
Qy	420	EMRATGPASTHSISTNAANQDVV	442	
Db	447	DQRPIPGIRMYQIQYNSASDAIL	469	

Query Match	3.5%; Score 93; DB 6; Length 553;
Best Local Similarity	21.5%; Pred. No. 2.7;
Matches	117; Conservative 83; Mismatches 226; Indels 118; Gaps 27
Qy	46 RASEARLGAVIRAEARHYVGLTTGPGPLANRLISGENVRTLQANLVHHLASGVGPVLDW-- 103
Db	12 RALPRRAGLVRNA-----LGSFIAARNFHASNTHLQKTGTAEKSSILIERI 59
Qy	104 ----TTARAMVLARLVSAQASGSEGTTLARLLDLINSELAPAVP--SRGTVGASGDLTTP 158
Db	60 LGADTSDVLEETGRVLSIGDG-----IARVHGLRNVAQAEEMVEPSSGLKGMSLNLEP 111
Qy	159 LAHWVLC-----LQCRGDFLDRDTRLD---GAGRLRGRLOPLDLSHRDALALVNGTSA 210
Db	112 DNVGVVFGNDKLTKEGDI VKRTGAI VDVVPVGBELL--GRVV-----DALG-----NA 157
Qy	211 MTGIALVNAHACRHGLGNWALVALLAECLRGRTAEWAAALSDLRP--HPGOKDAARLRA 269
Db	158 IDKGPIGSKTRREVGLKAPGIIIPRIS--VREPWTGIIKAVDSLVPIGRGQRELI--IGD 213
Qy	270 RVDSGARV--VRHVTAERRLDAGDIGTEPEAGODAYSRLCAPQVILGAGPDTLAWHDRVL-- 326
Db	214 ROTGKTSIAITTIINQKRFNDG-----SDEKKLYCIYVAIGKRSYVAQLVKRLTD 265
Qy	327 -----TIELNAVTDNVPFPDGSVPALHGGNFMGQ-----HVALTSDALATAVTVLA 373
Db	266 ADAMKYTIIVSA--TASDAAPLOYLAP--YSGCSNGEYFRDNGKHALIYYDDLSK-----Q 317
Qy	374 GLAERQIARLTDERLNRGLPP-----FLH-----RGPAGLNGSFMGAQVTTATALL-----AEM 421
Db	318 AVAYRQMSLLRRPPGREAYPGDVFLHLSRLLERAAKMNDAPFGGSLTALPVIETQAGDV 377
Qy	422 RATGPASIHISITN-----AANQDVVSLGTIAARLCREKIDR--WAILATLALCL 470
Db	378 SAVIPTNVIISITDQIFLETETLFYKGRIPAINVLSVRSVGSAAQTRAMKQVAGTMKUEL 437
Qy	471 AQ-----AAELRCGSLDGVS-----PAGKGLVQALRE--QFPLETDRLPGQBIATAATHLL 521
Db	438 AQYREVAAPFAQFGSLDAAATQQLSRGVRUTELLKQGYSPMAITEEQVAVIYAGVRGYLD 497
Qy	522 QQSP 525
Db	498 KLEP 501

RESULT 15
US-10-925-970-3
; Sequence 3, Application US/10925970
; Publication No. US20050249741A1

; GENERAL INFORMATION:
; APPLICANT: KIMURA, Toru
; APPLICANT: KIKUCHI, Kaoru
; TITLE OF INVENTION: NOVEL SEMAPHORIN GENE: SEMAPHORIN W
; FILE REFERENCE: 0020-4546P
; CURRENT APPLICATION NUMBER: US/10/925,970
; CURRENT FILING DATE: 2004-08-26
; PRIOR APPLICATION NUMBER: US/09/284,180
; PRIOR FILING DATE: 1999-06-09
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 776
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-10-925-970-3

Query Match 3.5%; Score 93; DB 6; Length 776;
Best Local Similarity 26.0%; Pred. No. 4.3;
Matches 75; Conservative 34; Mismatches 106; Indels 74; Gaps 17;
QY 233 TALLAECL-----RGRTEAWAALSDLRPHPG-----OKDAAA----- 265
Db 294 TFLKADLLCPGPEHGRASGVQLQAWAELRPPQGAGTPIFYGIFSSQWEGAAISAVCAFRPQ 353
QY 266 RLRRVDSGARVVRHVIAERRLDAGDGT-EPEAGQDAYSLRCAPOVLGAGFTTLAWHDR 324
Db 354 DIRAVLNGPPRELKH-DCNRGLPMDNEVPQPRPG-ECIANNMKLOQFGS---SLSLPDR 408
QY 325 VLT-IELNAVTDNPVPPDGSGVPALHGGNFMGOHVALTSD-----ALATAVTVLAGLAER 378
Db 409 VLTFFIRDHPLMDFVFPAD-----GRPLLVTDTAYLRVVVHRVTSLSG-KEY 455
QY 379 QIARLTDERLNRGLPPLHHRGPAGLNSGFMGAQVTTATALLAEMRATCPAS-----IHSIST 434
Db 456 DVLYLGTE-----DGHLLHRAVR-----IGAQLSVLEDLALPPEQPQVPMKLYHDWLL 503
QY 435 NAANQDVVSLGTI-AAAL--CREKIDRWAEILA---ILALCLAAQAAELR 477
Db 504 VGSHTVTVQVNTSNCGRQLQSCSICILAQDPVCAWSFRLDACVAHAGEHR 552

Search completed: December 4, 2005, 09:16:22
Job time : 13 secs